

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ (государственный университет)
ФАКУЛЬТЕТ УПРАВЛЕНИЯ И ПРИКЛАДНОЙ МАТЕМАТИКИ
КАФЕДРА «ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ»

Стенина Мария Михайловна

**Согласование прогнозов в задачах прогнозирования
иерархических временных рядов**

010990 — Интеллектуальный анализ данных

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА МАГИСТРА

Научный руководитель:

д. ф.-м. н. Стрижов Вадим Викторович

Москва

2015

Содержание

Введение	5
Обозначения	9
1 Постановка задачи согласования прогнозов	10
1.1 Иерархические временные ряды	10
1.2 Постановка задачи согласования прогнозов	13
2 Обзор существующих алгоритмов согласования прогнозов	15
2.1 Методы, использующие прогнозы части временных рядов	15
2.2 Методы, использующие независимые прогнозы всех временных рядов .	18
3 Алгоритм согласования прогнозов GTOp	21
3.1 Анализ недостатков существующих алгоритмов согласования прогнозов	21
3.2 Задача согласования прогнозов как поиск равновесия Нэша в антагонистической игре	23
3.2.1 Понятие антагонистической игры.	24
3.2.2 Антагонистическая игра, описывающая задачу согласования прогнозов.	25
3.2.3 Существование равновесия Нэша.	27
3.3 Алгоритм согласования прогнозов GTOp	31
3.4 Функции потерь, применимые для GTOp	32
3.5 Модификация GTOp с учетом надежности независимых прогнозов — MGTOp	34
4 Вычислительный эксперимент	36
4.1 Экспериментальные данные	36
4.1.1 Посуточная загруженность узлов РЖД	36
4.1.2 Почасовое потребление электроэнергии	39
4.2 Демонстрация свойств алгоритма GTOp	41
4.2.1 Оптимизационная задача	41
4.2.2 Результаты	41

4.3	Выбор весов функции потерь для MGTOp	42
4.3.1	Оптимизационная задача	42
4.3.2	Результаты	43
	Заключение	46
	Список литературы	47

Аннотация

Рассмотрена задача одновременного прогнозирования набора временных рядов, объединенных в иерархическую многоуровневую структуру. Требуется, чтобы полученные прогнозы удовлетворяли физическим ограничениям и структуре иерархии. Предлагается алгоритм согласования прогнозов иерархических временных рядов GTOp (Game-theoretically optimal reconciliation), гарантирующий неухудшение качества прогнозов после проведения процедуры согласования по сравнению с качеством прогнозов, полученных для каждого временного ряда независимо. Подход базируется на поиске равновесия Нэша в антагонистической игре заданного вида и сводит задачу согласования прогнозов к задаче оптимизации с ограничениями типа равенства и неравенства. Также предложена модификация алгоритма согласования GTOp, позволяющая учитывать надежность независимых прогнозов путем минимальной коррекции наиболее надежных прогнозов в ходе процедуры согласования. Работа алгоритма согласования демонстрируется на данных о посуточной загрузке узлов РЖД и на данных почасового потребления электроэнергии. Проведено сравнение качества согласованных прогнозов, полученных алгоритмом GTOp, с качеством прогнозов, полученных с помощью существующих алгоритмов согласования. Демонстрируется, что алгоритм GTOp позволяет получать согласованные прогнозы более высокого качества, чем существующие алгоритмы согласования.

Ключевые слова: *иерархические временные ряды, прогнозирование, согласование прогнозов временных рядов, антагонистическая игра, равновесие Нэша.*

Введение

Актуальность темы. Магистерская работа посвящена исследованию задачи прогнозирования иерархических временных рядов. Временные ряды связаны в иерархическую многоуровневую структуру, в которой ряды каждого следующего более высокого уровня формируются путем поэлементного суммирования некоторой части (возможно, всех) временных рядов предыдущего уровня. Иерархическую структуру можно описать с помощью графа, который в простейшем случае является деревом. По имеющимся историческим значениям временных рядов и информации об иерархической структуре необходимо построить прогноз каждого временного ряда для следующего отсчета времени. Полученные прогнозы должны удовлетворять структуре иерархии и физическим ограничениям (если они заданы) и минимизировать заданную функцию потерь.

Решение поставленной задачи актуально для областей энергетики [1], туризма [2], торговли [3], железнодорожных перевозок [4]. Во всех перечисленных областях наблюдаемые детализированные временные ряды возможно агрегировать по одному или нескольким параметрам, в результате чего возникает иерархическая структура, описывающая связь между временными рядами.

Решение поставленной задачи включает два этапа. Первый — получение независимых прогнозов для всех временных рядов иерархии, второй — проведение процедуры согласования прогнозов. Алгоритмы, которые используются для решения задач на каждом из этапов, влияют не только на качество прогнозирования, но и накладывают ограничения на алгоритмы, которые можно использовать на другом этапе. Также на используемые алгоритмы накладывают ограничения структура иерархии, функция потерь, физические ограничения на прогнозы.

Прогнозы, получаемые для каждого временного ряда независимо от других не будут согласованы со структурой иерархии, что будет приводить к противоречиям при их анализе. С помощью корректировки (согласования со структурой иерархии) прогнозов можно избежать противоречий, а также повысить качество прогнозирования, если процедура согласования учитывает заданную функцию потерь. В настоящее время существуют несколько семейств алгоритмов согласования прогнозов, у каждого из них есть достоинства и ограничения, сужающие множество задач, в

которых применимы алгоритмы из каждого семейства.

При восходящем согласовании [5] используются только прогнозы временных рядов нижнего уровня иерархии. Такой способ согласования с успехом может применяться в задачах с неплоской иерархической структурой, когда граф иерархии не является деревом, так как не приводит к противоречиям в вершинах графа, которые имеют более одной родительской вершины. Однако алгоритмы из этого семейства не учитывают особенностей заданной функции потерь и могут привести к ухудшению качества прогнозирования, если временные ряды нижнего уровня сильно зашумлены. При нисходящем согласовании [5] используются прогнозы временного ряда верхнего уровня иерархии и пропорции, в которых значения временных рядов верхних уровней делятся между значениями рядов более низких уровней. Алгоритмы этого семейства могут применяться при зашумленных рядах нижнего уровня, но не применимы для неплоских иерархий из-за возможных коллизий в вершинах с несколькими родительскими. Невозможность учесть особенности функции потерь является еще одним недостатком этого семейства алгоритмов. Регрессионный подход к согласованию [2, 6] использует прогнозы всех временных рядов и может применяться при неплоской иерархической структуре. Однако этот подход использует предположение о несмещенности прогнозов, что накладывает сильные ограничения на алгоритмы, которые используются для получения независимых прогнозов. Еще один недостаток этого подхода — необходимость оценки ковариационной матрицы шума, что часто бывает невозможно. Ни один из описанных выше подходов не гарантирует улучшения качества прогноза после проведения процедуры согласования.

Таким образом среди существующих алгоритмов согласования нет такого, который гарантировал бы улучшение качества прогнозов после проведения процедуры согласования, был бы применим для иерархических структур любого вида, позволял бы использовать при прогнозировании различные функции потерь и не накладывал бы строгие ограничения на алгоритмы получения независимых прогнозов на первом этапе решения поставленной задачи.

Цель работы. Целью работы является разработка метода согласования прогнозов иерархических временных рядов, который

- применим для иерархических структур любого вида;
- не накладывает на алгоритмы получения независимых прогнозов временных рядов существенных ограничений;
- гарантирует неухудшение качества прогнозов после проведения процедуры согласования прогнозов со структурой иерархии;

теоретическое обоснование свойств разработанного метода прогнозирования, а также демонстрация этих свойств на практике.

Методы исследования. Для достижения поставленных целей используется аппарат теории игр [7–9], теории выпуклой оптимизации [10] и теории вероятностей. Для программной реализации разработанного алгоритма согласования прогнозов использовалась среда MATLAB.

Основные положения, выносимые на защиту.

1. Алгоритм согласования прогнозов GTOp (Game-theoretically optimal reconciliation), применимый для иерархических структур любого вида, накладывающий на алгоритмы получения независимых прогнозов только физические ограничения и гарантирующий неухудшение качества прогнозов после проведения процедуры согласования.
2. Модификация алгоритма GTOp, позволяющая учитывать надежность прогнозов различных временных рядов в ходе процедуры их согласования.
3. Решение прикладной задачи согласования прогнозов объемов погрузки набора грузов в заданных узлах РЖД.
4. Решение прикладной задачи согласования прогнозов объемов потребления электроэнергии.
5. Экспериментальное исследование разработанных алгоритмов, содержащее их сравнение с аналогами.

Научная новизна. Разработан новый подход к задаче согласования прогнозов иерархических временных рядов. Обнаружено семейство функций потерь, для которых может применяться этот подход. Предложен алгоритм согласования прогнозов с учетом надежности независимых прогнозов для каждого временного ряда в иерархической структуре.

Теоретическая значимость. Данное исследование вносит существенный вклад в область согласования прогнозов иерархических временных рядов, связанный с разработкой нового подхода к согласованию прогнозов, который не имеет недостатков, присущих ранее разработанным методам, и, следовательно, имеет более широкую область применения.

Практическая значимость. Разработанные методы показывают существенное улучшение качества решения реальных задач прогнозирования иерархических временных рядов. Разработан программный модуль, реализующий алгоритм согласования прогнозов GTOp и его модификацию.

Степень достоверности и апробация работы. Достоверность результатов подтверждена математическими доказательствами, экспериментальной проверкой полученных методов на реальных задачах прогнозирования временных рядов, публикациями результатов исследования в рецензируемых научных изданиях, в том числе рекомендованными ВАК. Результаты работы докладывались и обсуждались на следующих научных конференциях

- 10-я Международная конференция «Интеллектуализация обработки информации», 2014 г.;
- 57 научная конференция МФТИ с международным участием, 2014 г. [11];
- XXII Международная научная конференция студентов, аспирантов и молодых ученых «Ломоносов-2015», 2015 г. [12].

Публикации по теме. Основные результаты по теме магистерской работы изложены в изданиях из списка ВАК [4, 13–17], двух сборниках докладов конференций [11, 12] и других печатных изданиях [18, 19].

Обозначения

- Количество временных рядов в иерархии — d ;
- длина истории — T ;
- матрица временных рядов — X , $d \times T$;
- срез иерархии (столбец матрицы X) — χ , $d \times 1$;
- матрица связи иерархии — S ;
- функция потерь для иерархии — l_h ;
- множество векторов, удовлетворяющих структуре иерархии, — \mathcal{A} ;
- множество векторов, удовлетворяющих физическим ограничениям, — \mathcal{B} ;
- вектор независимых прогнозов — $\hat{\chi}$;
- вектор согласованных прогнозов — $\hat{\varphi}$.

1 Постановка задачи согласования прогнозов

1.1 Иерархические временные ряды

Набор временных рядов, взаимосвязи между которыми можно задать с помощью многодольного графа, называются иерархическими временными рядами. Связь между рядами задается параметром (размерностью), по которому происходит агрегирование информации. Такими параметрами могут выступать время, географическое разбиение, группы грузов или товаров, подразделения компании, сегменты рынка и т.д. По агрегирующим параметрам иерархии можно разделить на два типа: с отношением иерархии на шкале значений и на шкале времени.

Синхронные (contemporaneous) иерархические временные ряды. Ряды с агрегированием по значениям. Среди параметров агрегирования нет времени. Каждая вершина графа иерархии соответствует одному временному ряду, и в любой момент времени значение временного ряда в вершине равно сумме значений временных рядов в потомках этой вершины. В простейшем случае граф иерархии является деревом, и у каждой вершины есть только одна родительская вершина. Агрегирование в этом случае происходит в рамках одной размерности. В более сложных случаях агрегирование может проводиться по нескольким размерностям, и в этом случае граф иерархии деревом не является, и вершина в графе может иметь несколько родительских.

Периодические (temporal) иерархические временные ряды. Ряды с иерархической шкалой времени. Временные ряды агрегируются по периоду времени. Чем выше временной ряд находится в графе иерархии, тем большему периоду времени отвечают его значения.

Далее будут рассматриваться только синхронные иерархические временные ряды. На рис. 1 изображен пример простейшей иерархической структуры, в которой агрегирование осуществляется по одному параметру.

Будем обозначать временной ряд через вектор \mathbf{x} , элементы временного ряда будем снабжать индексом t , $t = 1, \dots, T$, где T — длина истории.

$$\mathbf{x} = \{x_t\}_{t=1}^T.$$

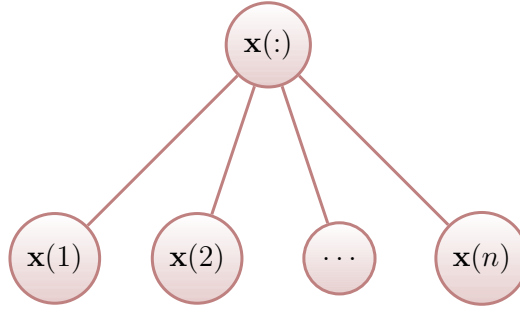


Рис. 1: Пример иерархических временных рядов

Общее количество временных рядов во всей иерархии будем обозначать d . Далее для наглядности будем рассматривать иерархию рядов, изображенную на рис. 1. Она содержит один временной ряд на верхнем уровне и n рядов на нижнем. Для этой иерархии $d = 1 + n$.

Временные ряды нижнего уровня будут обозначаться $\mathbf{x}(i)$, $i = 1, \dots, n$, где n — количество временных рядов на нижнем уровне. Временной ряд верхнего уровня обозначается как $\mathbf{x}(:)$. Во избежание путаницы условимся использовать нижние индексы для обозначения отсчетов времени, а индексы в скобках использовать для обозначения положения временных рядов в структуре иерархии. Элементы временных рядов, составляющих плоскую двухуровневую иерархию, обозначаются соответственно $x_t(i)$, $i = 1, \dots, n$, $x_t(:)$. Их соотношение задается формулой

$$x_t(:) = \sum_{i=1}^n x_t(i), \quad t = 1, \dots, T. \quad (1)$$

Будем называть соотношение (1) *условием согласованности*. Прогнозы этих временных рядов будем обозначать «шляпками», опуская нижние индексы, чтобы избежать излишне громоздких обозначений. Прогнозироваться будет всегда $(T + 1)$ -е значение временного ряда

$$\hat{x}(i), \quad i = 1, \dots, n, \quad \hat{x}(:).$$

Согласованные прогнозы будут также обозначаться без нижних индексов

$$\hat{y}(i), \quad i = 1, \dots, n, \quad \hat{y}(:) = \sum_{i=1}^n \hat{y}(i).$$

Запишем все временные ряды в матрицу, каждая строка которой соответствует одному временному ряду. Для иерархии с рис. 1 эта матрица будет размера $(1 + n) \times T$

и выглядеть следующим образом

$$X = \begin{pmatrix} x_1(:) & x_2(:) & \dots & x_T(:) \\ x_1(1) & x_2(1) & \dots & x_T(1) \\ \dots & \dots & \dots & \dots \\ x_1(n) & x_2(n) & \dots & x_T(n) \end{pmatrix}. \quad (2)$$

Будем называть срезом иерархии в момент времени t столбец матрицы X , соответствующий моменту времени t . Для удобства записи введем векторы, соответствующие срезу иерархии в момент времени t , прогнозам и согласованным прогнозам. В этих векторах значения, соответствующие разным временным рядам, записаны в столбец, начиная с верхнего уровня иерархии и заканчивая нижним уровнем

$$\chi_t = \begin{pmatrix} x_t(:) \\ x_t(1) \\ \dots \\ x_t(n) \end{pmatrix}, \quad \hat{\chi} = \begin{pmatrix} \hat{x}(:) \\ \hat{x}(1) \\ \dots \\ \hat{x}(n) \end{pmatrix}, \quad \hat{\varphi} = \begin{pmatrix} \hat{y}(:) \\ \hat{y}(1) \\ \dots \\ \hat{y}(n) \end{pmatrix}.$$

Условие (1) для векторов χ_t , $\hat{\chi}$, $\hat{\varphi}$ запишем, введя матрицу связей размера $1 \times (n+1)$

$$S = \begin{pmatrix} -1 & 1 & \dots & 1 \end{pmatrix}.$$

Тогда условие согласованности запишется кратко

$$S\chi_t = 0, \quad S\hat{\varphi} = 0.$$

В случае, когда иерархия имеет более сложную структуру, чем на рис. 1, векторы χ_t , $\hat{\chi}$, $\hat{\varphi}$ имеют размерность d , матрица X имеет ровно d строк, временные ряды в ней записываются от верхних уровней к нижним. А размерность матрицы связей S равна $c \times d$, где c — количество узлов в графе иерархии или, другими словами, количество ограничений-равенств, наложенных на элементы срезов иерархии χ_t .

Определим множество

$$\mathcal{A} = \{\chi \in \mathbb{R}^d \mid S\chi = 0\}, \quad (3)$$

где χ — произвольный d -мерный вектор, а S — заданная матрица связей. Отметим, что все срезы χ_t , $t = 1, \dots, T$, лежат в множестве \mathcal{A} . Также в нем должны лежать согласованные прогнозы $\hat{\varphi}$.

В ряде задач прогнозы должны удовлетворять некоторым ограничениям, связанным с физической природой прогнозируемой величины. В связи с чем введем множество

$$\mathcal{B} = \{\chi \in \mathbb{R}^d \mid \chi(i) \in [A_i, B_i], A_i, B_i \in [-\infty, +\infty], i = 1, \dots, d\}, \quad (4)$$

где χ — произвольный d -мерный вектор, A_i, B_i задают отрезок, в котором должна находиться i -тая компонента этого вектора. Например, в случае $A_1 = \dots = A_d = 0, B_1 = \dots = B_d = +\infty$ вектор χ лежит в положительном октанте. А конечные значения $A_1, \dots, A_d, B_1, \dots, B_d$ могут задавать интервалы, в которых должны лежать прогнозы. Отсутствие каких-либо ограничений задается значениями $A_1 = \dots = A_d = -\infty, B_1 = \dots = B_d = +\infty$.

1.2 Постановка задачи согласования прогнозов

Пусть дан набор из d временных рядов, значения которых записаны в матрицу X размера $d \times T$

$$X = \begin{pmatrix} x_1(1) & x_2(1) & \dots & x_T(1) \\ x_1(2) & x_2(2) & \dots & x_T(2) \\ \dots & \dots & \dots & \dots \\ x_1(d) & x_2(d) & \dots & x_T(d) \end{pmatrix},$$

где каждый столбец χ_t соответствует срезу в момент времени t , а каждая строка \mathbf{x}_i — одному временному ряду. Пусть структура иерархии задана матрицей связи S так, что для всех $t = 1, \dots, T$ выполнено условие согласованности

$$S\chi_t = 0.$$

Пусть даны прогнозы значений $\hat{\chi}$ для всех временных рядов в момент времени $T + 1$ и задана функция суммарных потерь при прогнозировании иерархии

$$l_h(\chi_{T+1}, \hat{\chi}). \quad (5)$$

Введя все необходимые дополнительные обозначения, можно сформулировать задачу поиска согласованных прогнозов. Требуется найти вектор прогнозов $\hat{\varphi}$, удовлетворяющий следующим требованиям.

Согласованность: вектор прогнозов $\hat{\varphi}$ должен удовлетворять структуре иерархии, заданной матрицей связи S , то есть $\hat{\varphi} \in \mathcal{A}$.

Ограничения: вектор прогнозов $\hat{\varphi}$ должен удовлетворять наложенным ограничениям, то есть $\hat{\varphi} \in \mathcal{B}$.

Качество: общие потери при использовании согласованных прогнозов не должны превышать общие потери при использовании независимых прогнозов, то есть $l_h(\chi_{T+1}, \hat{\varphi}) \leq l_h(\chi_{T+1}, \hat{\chi})$.

2 Обзор существующих алгоритмов согласования прогнозов

В рамках этого обзора будут рассмотрены только синхронные иерархические временные ряды. Основное свойство, которым обладают временные ряды, объединенные в иерархическую структуру такого типа, — временной ряд, соответствующий вершине графа иерархии, получен поэлементным суммированием временных рядов, соответствующих потомкам этой вершины. Наиболее детализированные временные ряды называются временными рядами нижнего уровня, наименее детализированные (то есть наиболее агрегированные) — рядами верхнего уровня, остальные — рядами промежуточных уровней. При прогнозировании синхронных иерархических временных рядов необходимо, чтобы полученные прогнозы также обладали описанным свойством, то есть были *согласованы*. Выполнения этого свойства для прогнозов можно добиться разными путями. Первый подход заключается в получении прогнозов только для некоторой части временных рядов из имеющегося набора. Прогнозы остальных рядов получаются путем использования взаимосвязей в графе иерархии и, если необходимо, соотношений значений разных временных рядов, наблюдавшихся в истории. Второй подход заключается в получении прогнозов для всех временных рядов независимо. Очевидно, что это приводит к нежелательным последствиям несовпадения прогнозов более высоких уровней иерархии с суммой прогнозов более низких уровней. Следовательно, вторым шагом в рамках этого подхода является применение к полученным независимым прогнозам процедуры согласования.

Далее работа содержит более подробное описание различных способов получения согласованных прогнозов для иерархических временных рядов. Сперва будут описаны методы, использующие прогнозы части рядов для получения всех остальных прогнозов. Далее будут рассмотрены методы, использующие независимые прогнозы всех временных рядов и подвергающие их процедуре согласования.

2.1 Методы, использующие прогнозы части временных рядов

Существующие методы прогнозирования иерархических временных рядов в рамках этого подхода обычно включают восходящие и нисходящие методы или их комбина-

цию. В нисходящем методе прогнозируются агрегированные временные ряды, и затем прогнозы деагрегируются согласно пропорциям, наблюдающимся в истории. В [20] обсуждаются несколько способов выбора пропорций. При использовании восходящего метода сперва прогнозируются все неагрегированные временные ряды нижнего уровня иерархии, а затем используется простое агрегирование для получения прогнозов более высоких уровней иерархии. На практике эти два метода часто комбинируются (восходяще-нисходящий подход), при этом строятся прогнозы временных рядов одного из промежуточных уровней иерархии, затем их агрегируют, чтобы получить прогнозы более высоких уровней и деагрегируют для получения прогнозов более низких уровней. Ни один из этих методов не рассматривает структуру иерархии в целом.

Восходящий подход применим для иерархических структур, в которых агрегирование проводится как по одному, так и по нескольким параметрам. Нисходящий и восходяще-нисходящий подходы применимы только в случаях, когда граф иерархии является деревом. В противном случае они могут привести к противоречивым результатам в вершинах графа, имеющих более одной родительской вершины. Главное достоинство этих методов заключается в их простоте и интерпретируемости. Недостаток заключается в том, что они рассматривают только небольшую часть иерархической структуры, а не весь граф целиком.

Основная часть литературы в этой области посвящена сравнительному анализу восходящих и нисходящих методов. В одной из самых ранних работ [21] авторы утверждают, что неагрегированные данные содержат большое количество ошибок, поэтому нисходящий подход дает более точные прогнозы. К похожим выводам приходят в [22, 23]. В статье [24] также обсуждается, что прогнозирование агрегированных временных рядов более предпочтительно. С другой стороны, в работах [25, 26] представлена точка зрения, что основные потери информации происходят при агрегировании, и поэтому восходящий подход обеспечивает большую точность прогнозов. Авторы статьи [5] сравнивают оба метода и приходят к выводу, что восходящий подход предпочтительнее при выполнении ряда условий для структуры иерархии и горизонта прогноза. В [27] исследуются устойчивость и смещение прогнозов, полученных обоими методами, и заключается, что восходящий подход надежнее, кроме

случаев наличия пропусков и выбросов в данных нижнего уровня иерархии.

Практические исследования подтверждают преимущество восходящего подхода над нисходящим. К примеру, в статье [28] обнаружено, что использование для построения прогнозов данных о заработной плате в разных сегментах рынка приводит к получению более точных прогнозов, чем использование агрегированных данных по фирмам. В [29] проведено сравнение сегментированной эконометрической модели с агрегированной, и обнаружено, что сегментированная модель обеспечивает более точны прогнозы продаж и прибыли. Исследование спроса на телефоны в [30] также демонстрирует, что построение детализированной модели обеспечивает нахождение более точных прогнозов. В [31] использовались данные о темпах роста валового внутреннего продукта в 18 странах, и тоже детализированная модель дала более точные прогнозы. В [32] были сгенерированы синтетические данные с двухуровневой иерархической структурой и двумя временными рядами на нижнем уровне. Проведенные в этой работе исследования еще раз подтвердили, что на практике восходящий подход дает более точные прогнозы, чем нисходящий, особенно, если временные ряды нижнего уровня иерархии сильно коррелированы.

В эконометрической литературе также встречается метод повышения точности прогнозов путем агрегирования компонент прогнозов вместо использования прогнозов агрегированных временных рядов [31, 33–36].

Что касается теоретических исследований, в [37, 38] приводятся теоретические аргументы, показывающие, что эффективность агрегирования зависит от ковариационной структуры временных рядов, входящих в иерархию. Автор работы [39] рассматривает несколько моделей временных рядов и демонстрирует, что нет однозначного превосходства одного метода над другим. В статье [40] заключается, что все эвристические подходы к прогнозированию иерархических временных рядов применимы на практике в разных задачах. В [41] предлагается использовать одновременно оба метода, но никаких новых подходов не предлагается. В [42] описываются все известные на тот момент подходы к прогнозированию иерархических временных рядов, однако не предлагается новых.

В настоящее время предлагаются различные модификации восходящего и нисходящего подхода с привлечением различных инструментов машинного обучения. В

работах [43, 44] предлагается вариант восходящего прогнозирования, в котором при агрегировании прогнозов нижнего уровня используется взвешенная сумма, веса которой настраиваются и обновляются в режиме он-лайн. В [45] описан вариант выбора пропорций при нисходящем прогнозировании с помощью вариационного байесовского вывода и вариационного ЕМ-алгоритма.

2.2 Методы, использующие независимые прогнозы всех временных рядов

Методы этого семейства являются более сложными, чем описанные ранее, однако использование информации о всех независимых прогнозах дает возможность неявно учитывать взаимосвязи, которые, возможно, есть между временными рядами в иерархии, но не отражены в структуре графа.

В работе [46] предлагается способ согласования полученных независимо друг от друга прогнозов. Авторы используют линейную комбинацию прогнозов, полученных для временного ряда в некоторой вершине и временных рядов в потомках этой вершины. В статье рассмотрены два варианта выбора весов линейной комбинации. Утверждается, что предложенный способ согласования дает более точные прогнозы, нежели независимо полученные прогнозы, в случае, когда временные ряды нижнего уровня описывают нерегулярные и сильно изменчивые процессы. Этот метод применим только для иерархий с древесной структурой, в случае наличия у вершины более одного родителя он приведет к противоречиям.

Еще одно направление исследований в этой области — создание алгоритма, который в автоматическом режиме выбирал бы модель согласования прогнозов в каждом узле иерархии. В [47] утверждается, что описание иерархической структуры одной моделью согласования является очень грубым, и предлагается алгоритм, который итеративно в каждом узле иерархии настраивает свою модель согласования, которая может быть взвешенным вариантом нисходящего или восходящего метода либо линейной комбинацией прогнозов двух уровней, как предложено в работе [46].

В статье [2] предлагается обобщенная модель согласования прогнозов иерархических временных рядов. Нисходящий, восходящий и нисходяще-восходящий методы являются частными случаями этой модели. Авторами впервые предложено разбить

задачу получения согласованных прогнозов на два этапа. На первом этапе для каждого временного ряда строится прогноз при помощи какого-либо алгоритма прогнозирования. На втором этапе полученные прогнозы подвергаются процедуре согласования. Предполагается, что вектор независимых прогнозов линейно зависит от прогнозов временных рядов нижнего уровня с некоторой шумовой добавкой. Предполагается, что шум имеет нулевое математическое ожидание, то есть независимые прогнозы являются несмещенными оценками. Также предполагается, что шумовые слагаемые удовлетворяют тем же условиям согласованности, которым должны удовлетворять прогнозы. С учетом всех этих предположений решается задача линейной регрессии для получения новой оценки прогнозов временных рядов нижнего уровня, которая является несмещенной и имеет минимально возможную дисперсию. Согласованные прогнозы получаются путем суммирования полученных оценок на нижнем уровне иерархии. Подход, предложенный в этой статье, ранее упоминался в работах по экономике [48–53]. Авторами в [54] презентуется пакет на языке программирования R, реализующий предложенный алгоритм получения согласованных прогнозов. В статье [2] приводится описание алгоритма, примененного для иерархических структур, графы которых являются деревьями. Однако этот алгоритм легко обобщается на случай, когда иерархия описывается не деревом, что и сделано этими же авторами в [6].

Статья [4] посвящена алгоритму поэтапного согласования прогнозов в каждом узле иерархии. Для получения согласованных прогнозов для каждого узла решается оптимизационная задача с ограничениями типа равенства и неравенства. Ограничения типа неравенства задаются доверительными интервалами, которые оцениваются для независимых прогнозов. Поэтому данный алгоритм применим только для задач, в которых возможно корректно оценивать доверительные интервалы независимых прогнозов. Согласование ведется сверху вниз по иерархии, при этом оптимизационная задача решается для каждого узла отдельно. Основная идея предложенного алгоритма заключается в том, что, чем больше доверительный интервал для прогноза, тем сильнее его можно корректировать при проведении процедуры согласования. Также в статье сравнивается качество согласованных прогнозов, полученных предлагаемым авторами алгоритмом, и прогнозов, полученных с помощью алгорит-

ма из [2]. Показано, что предлагаемый авторами алгоритм не уступает алгоритму с использованием регрессионной модели. Оба алгоритма тестировались на данных, описывающих посуточную загруженность железнодорожных узлов.

3 Алгоритм согласования прогнозов GTOp

В данном разделе проведен анализ наиболее популярных алгоритмов согласования прогнозов, а именно восходящего и нисходящего подхода, а также регрессионного подхода [6, 54]. Описаны основные ограничения и предположения, на которых основаны эти алгоритмы согласования. Обнаруженные ограничения разделены на две категории:

1. принципиальные ограничения, входящие в постановку задачи согласования прогнозов, от которых нельзя отказаться;
2. ограничения, не связанные с постановкой задачи согласования прогнозов, от которых можно отказаться при разработке нового алгоритма согласования прогнозов.

Задача согласования прогнозов переформулирована в терминах антагонистической игры так, что включает ограничения только первой категории. Доказывается, что для полученной игры при выполнении ряда предположений о свойствах функции потерь существует равновесие Нэша, которое определяет вектор согласованных прогнозов и способ его вычисления, удовлетворяющий всем требованиям, сформулированным при постановке задачи согласования прогнозов. Затем представлен обзор семейства функций потерь, для которых выполняются необходимые предположения. В последней части этого раздела описана модификация полученного алгоритма согласования прогнозов, которая позволяет учесть надежность имеющихся независимых прогнозов. Модифицированный алгоритм проводит процедуру согласования с минимальной коррекцией наиболее надежных независимых прогнозов.

3.1 Анализ недостатков существующих алгоритмов согласования прогнозов

При восходящем согласовании [5] используются только прогнозы временных рядов нижнего уровня иерархии, о которых не делается каких-либо предположений. Такой способ согласования с успехом может применяться в задачах с неплоской иерархической структурой, когда граф иерархии не является деревом, так как не приводит

к противоречиям в вершинах графа, которые имеют более одной родительской вершины. Однако алгоритмы из этого семейства не учитывают особенностей заданной функции потерь и могут привести к ухудшению качества прогнозирования, если временные ряды нижнего уровня сильно зашумлены.

При нисходящем согласовании [5] используются прогнозы временного ряда верхнего уровня иерархии и пропорции, в которых значения временных рядов верхних уровней делятся между значениями рядов более низких уровней. Ограничений на свойства прогнозов тоже не накладывается. Алгоритмы этого семейства могут применяться при зашумленных рядах нижнего уровня, но не применимы для неплоских иерархий из-за возможных коллизий в вершинах с несколькими родительскими. Невозможность учесть особенности функции потерь является еще одним недостатком этого семейства алгоритмов.

Нисходящий и восходящий подходы также не учитывают физических ограничений на прогнозы временных рядов.

Регрессионный подход к согласованию [2,6] использует прогнозы всех временных рядов и может применяться при неплоской иерархической структуре. Однако этот алгоритм согласования накладывает на полученные независимые прогнозы ряд ограничений, которые подробно рассматриваются в работе [55], где высказывается критика в адрес регрессионного подхода к согласованию прогнозов иерархических временных рядов. Авторы замечают, что предположение о несмещенности независимых прогнозов ограничивает набор алгоритмов прогнозирования, которые могут быть использованы для их получения. Например, невозможно использовать LASSO [56], так как этот алгоритм дает смещенные оценки. Однако основная трудность заключается в получении точной оценки ковариационной матрицы шума для каждого момента времени, что на практике часто оказывается невыполнимым. Также предположение о согласованности шумовых слагаемых является очень строгим ограничением, которое сильно ограничивает применимость такого алгоритма согласования прогнозов.

Авторы [55] предлагают алгоритм согласования прогнозов Game-theoretically Optimal reconciliation (GTOp), накладывающий на независимые прогнозы лишь ограничения, связанные с физическим смыслом прогнозируемых величин. Никаких предположений о вероятностной структуре временных рядов не делается. Это позволяет

использовать широкий спектр алгоритмов прогнозирования для получения независимых прогнозов. Далее задача согласования прогнозов рассматривается как поиск равновесия Нэша в антагонистической игре, в которой первый игрок выбирает согласованные прогнозы, а второй — действительные значения временных рядов, которые одновременно удовлетворяют условию согласованности и физическим ограничениям. Этот подход описан в книге [9]. Авторами доказывается, что при использовании в качестве функции потерь квадрата взвешенного евклидового расстояния в рассматриваемой игре существует равновесие Нэша, для нахождения которого необходимо решить оптимизационную задачу с ограничениями типа равенства и неравенства. Искомый вектор согласованных прогнозов совпадает с равновесной стратегией первого игрока. Доказывается, что согласованные прогнозы не менее точны, чем исходные независимые.

Однако ни регрессионный подход [2, 6], ни вариант алгоритма GTOp, предложенный в [55], не учитывают при согласовании прогнозов физические ограничения. В [2, 6] ограничения такого типа не рассматриваются в постановке задачи, а в [55] физические ограничения накладываются на независимые прогнозы, но не на согласованные. А также в этих работах в качестве функции потерь рассматривается только квадрат евклидового или взвешенного евклидового расстояния.

3.2 Задача согласования прогнозов как поиск равновесия Нэша в антагонистической игре

В этом подразделе задача согласования прогнозов, сформулированная выше, рассматривается как антагонистическая игра. Такое представление не влияет решение задачи согласования прогнозов и направлено лишь на достижение наглядности и интерпретируемости полученных результатов. Формулировка задачи согласования прогнозов в терминах теории игр была сделана так, чтобы исключить все недостатки существующих алгоритмов согласования прогнозов, описанные выше. Учтено, что функция потерь прогнозирования может быть произвольной, что на прогнозы могут быть наложены физические ограничения, которые должны выполняться как для независимых, так и для согласованных прогнозов. А также при формулировке антагонистической игры не сделано никаких предположений о несмещенности либо

других статистических свойствах независимых прогнозов.

В первой части подраздела приводятся необходимые определения и факты из теории игр [7–9], во второй части вводится антагонистическая игра, соответствующая задаче согласования прогнозов, в третьей части формулируется и доказывается теорема о существовании в этой игре равновесия Нэша, которая является обобщением теоремы, доказанной в [55], и приводится следствие из этой теоремы, однозначно определяющее выбор оптимального вектора согласованных прогнозов.

3.2.1 Понятие антагонистической игры.

Определение 1 Система

$$\Gamma = (\mathcal{M}, \mathcal{N}, L),$$

где \mathcal{M}, \mathcal{N} — непустые множества и функция $L: \mathcal{M} \times \mathcal{N} \rightarrow \mathbb{R}$, называется антагонистической игрой (игрой с нулевой суммой) в нормальной форме. Элементы $\mu \in \mathcal{M}, \nu \in \mathcal{N}$ называются стратегиями игроков 1 и 2 соответственно, L — функцией потерь игрока 1. Потери игрока 2 полагаются равными $-L(\mu, \nu)$.

Определение 2 Говорят, что игра разыгрывается в чистых стратегиях, если оба игрока из имеющихся наборов действий \mathcal{M}, \mathcal{N} выбирают по одному действию μ, ν соответственно.

Определение 3 Введем на множествах стратегий \mathcal{M}, \mathcal{N} вероятностные распределения $p(\mu)$ и $q(\nu)$ соответственно:

$$\int_{\mathcal{M}} p(\mu) d\mu = 1, \quad \int_{\mathcal{N}} q(\nu) d\nu = 1.$$

$p(\mu)$ и $q(\nu)$ задают смешанные стратегии в игре Γ , если игрок 1 выбирает действие в соответствии с распределением $p(\mu)$ и игрок 2 выбирает действие в соответствии с распределением $q(\nu)$.

Чистые стратегии являются частным случаем смешанных. Поэтому далее будут рассматриваться смешанные стратегии, за исключением специально оговоренных моментов. Обозначать стратегии будем $p(\cdot)$ и $q(\cdot)$.

Математическое ожидание потерь игрока 1 при паре смешанных стратегий $p(\cdot)$ и $q(\cdot)$ обозначим через

$$\bar{L}(p(\cdot), q(\cdot)) = \int_{\mathcal{M}} \int_{\mathcal{N}} p(\mu)q(\nu)L(\mu, \nu) d\mu d\nu.$$

Игрок 1 преследует цель минимизировать эту величину при любых действиях игрока 2.

Определение 4 Пара стратегий $(p(\cdot), q(\cdot))$ называется равновесием Нэша в смешанных стратегиях в игре Γ , если для любых $p(\cdot)'$, $q(\cdot)'$ выполнено неравенство

$$\bar{L}(p(\cdot), q(\cdot)') \leq \bar{L}(p(\cdot), q(\cdot)) \leq \bar{L}(p(\cdot)', q(\cdot)).$$

При равновесии Нэша ни одному из игроков не выгодно отклоняться от равновесной стратегии, если второй продолжает придерживаться равновесной стратегии. При этом игрок 1 минимизирует свои потери в ситуации, когда игрок 2 действует наиболее выгодным для себя образом. Отметим также, что равновесие Нэша является седловой точкой функции $\bar{L}(p(\cdot), q(\cdot))$.

Теорема 1 В антагонистической игре равновесие Нэша существует тогда и только тогда, когда определена величина

$$V = \min_{p(\cdot)'} \max_{q(\cdot)'} \bar{L}(p(\cdot)', q(\cdot)') = \max_{q(\cdot)'} \min_{p(\cdot)'} \bar{L}(p(\cdot)', q(\cdot)').$$

Величина V называется ценой игры. Доказательство этой теоремы в настоящей работе не приводится, при желании его можно найти в [7–9].

3.2.2 Антагонистическая игра, описывающая задачу согласования прогнозов.

Вернемся к рассмотрению введенных в разделе постановки задачи множеств \mathcal{A} (3) и \mathcal{B} (4). Напомним, что множество \mathcal{A} содержит все d -мерные векторы, удовлетворяющие структуре иерархии, заданной матрицей связи S . Множество \mathcal{B} содержит d -мерные векторы, удовлетворяющие ограничениям-неравенствам рассматриваемой задачи прогнозирования.

Будем рассматривать антагонистическую игру Γ , в которой игрок 1 выбирает вектор согласованных прогнозов $\hat{\varphi}$ из множества $\mathcal{A} \cap \mathcal{B}$, которые одновременно

удовлетворяют структуре иерархии, заданной матрицей связи S , и ограничениям-неравенствам, задающим множество \mathcal{B} , игрок 2 также выбирает вектор действительных значений χ_{T+1} , удовлетворяющих структуре иерархии и физическим ограничениям (можно считать игрока 2 природой).

Определим множества стратегий \mathcal{M} и \mathcal{N} игроков 1 и 2 как пересечение множеств $\mathcal{A} \cap \mathcal{B}$. Функцию потерь игрока 1 определим с помощью функции потерь по иерархии (5):

$$L(\hat{\varphi}, \chi_{T+1}) = l_h(\chi_{T+1}, \hat{\varphi}) - l_h(\chi_{T+1}, \hat{\chi}),$$

где вектор независимых прогнозов $\hat{\chi}$ считается заданным и не зависит от действий, выбираемых игроками. Считается, что $\hat{\chi} \in \mathcal{B}$. Такой выбор функции потерь игрока 1 связан с тем соображением, что при $L(\hat{\varphi}, \chi_{T+1}) = 0$ качество согласованных прогнозов не хуже, чем качество независимых прогнозов, а при $L(\hat{\varphi}, \chi_{T+1}) < 0$ и вовсе превосходит его.

Таким образом, получена антагонистическая игра

$$\Gamma = (\mathcal{A} \cap \mathcal{B}, \mathcal{A} \cap \mathcal{B}, l_h(\chi_{T+1}, \hat{\varphi}) - l_h(\chi_{T+1}, \hat{\chi})), \quad (6)$$

где игрок 1 выбирает вектор согласованных прогнозов $\hat{\varphi}$ из множества $\mathcal{A} \cap \mathcal{B}$ и игрок 2 выбирает вектор действительных значений элементов временных рядов χ_{T+1} из множества $\mathcal{A} \cap \mathcal{B}$. При этом первый игрок преследует цель минимизировать свои потери, выраженные функцией $L(\hat{\varphi}, \chi_{T+1})$, при любых действиях игрока 2, то есть при любом векторе действительных значений χ_{T+1} . Эта цель достигается, если игрок 1 воспользуется стратегией, входящей в равновесие Нэша.

В настоящее время подход, связанный с применением аппарата теории игр для анализа транспортно-экономических задач, активно используется. В частности, в работе [57] рассматривается модель конкурентного равновесия для анализа возможных проблем при формировании тарифной и инвестиционной политики управления железнодорожными грузоперевозками. В работе [58] исследуется один из классов транспортно-экономических задач, в которых поиск конкурентного равновесия сводится к поиску седловой точки в выпукло-вогнутой популяционной игре.

3.2.3 Существование равновесия Нэша.

В этой части раздела будет показано, что при выполнении ряда естественных требований к множествам \mathcal{A} (3) и \mathcal{B} (4) и функции суммарных потерь при прогнозировании иерархии l_h (5) в антагонистической игре (6), описывающей задачу согласования прогнозов, существует равновесие Нэша в чистых стратегиях. Также будет показано, что соответствующее этому равновесию значение функции потерь игрока 1 неположительно, что гарантирует неухудшение качества прогнозов при переходе от независимых прогнозов к согласованным. Рассмотрим эти требования.

Определение 5 Множество $\mathcal{C} \subseteq \mathbb{R}^d$ называется выпуклым [10], если для любых $\chi_1 \in \mathcal{C}$, $\chi_2 \in \mathcal{C}$ и любого $0 \leq \alpha \leq 1$ выполнено

$$\alpha\chi_1 + (1 - \alpha)\chi_2 \in \mathcal{C}.$$

Заметим, что множества \mathcal{A} (3) и \mathcal{B} (4) выпуклы и замкнуты.

Предположение 1 Будем предполагать, что пересечение множеств \mathcal{A} и \mathcal{B} не пусто $\mathcal{A} \cap \mathcal{B} \neq \emptyset$.

Требование непустого пересечения этих множеств естественно, так как в противном случае неразрешима задача поиска вектора согласованных прогнозов $\hat{\varphi}$, который должен одновременно принадлежать обоим множествам. Также отметим, что множество $\mathcal{A} \cap \mathcal{B}$ является выпуклым и замкнутым как пересечение двух выпуклых и замкнутых множеств [10].

Предположение 2 Будем считать, что функция суммарных потерь l_h (5) неотрицательна и равна нулю только при равенстве аргументов:

$$l_h(\chi_{T+1}, \hat{\chi}) \geq 0 \quad \text{для всех } \chi_{T+1}, \hat{\chi},$$

$$l_h(\chi_{T+1}, \hat{\chi}) = 0 \quad \Leftrightarrow \quad \chi_{T+1} = \hat{\chi}.$$

Равенство аргументов $\chi_{T+1} = \hat{\chi}$ соответствует случаю, когда прогноз полностью совпадает с действительными значениями. В этом случае потери равны нулю. Во всех остальных случаях потери при прогнозе положительные.

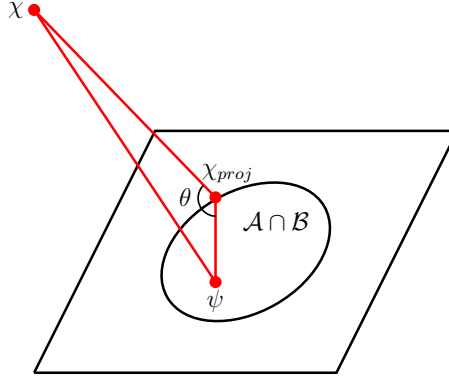


Рис. 2: Пояснение к предположению 4

Определение 6 Проекцией точки $\chi_0 \in \mathbb{R}^d$ на множество $\mathcal{C} \subseteq \mathbb{R}^d$, иницизированной функцией расстояния f называется точка

$$\chi_{proj} = \arg \min_{\chi \in \mathcal{C}} f(\chi, \chi_0).$$

Предположение 3 Пусть существует проекция точки из \mathbb{R}^d , соответствующей вектору независимых прогнозов $\hat{\chi}$, на выпуклое и замкнутое множество $\mathcal{A} \cap \mathcal{B}$, иницизированная функцией суммарных потерь l_h :

$$\chi_{proj} = \arg \min_{\chi \in \mathcal{A} \cap \mathcal{B}} l_h(\chi, \hat{\chi}).$$

Предположение 4 Пусть $\chi_{proj} = \arg \min_{\chi \in \mathcal{A} \cap \mathcal{B}} l_h(\chi, \hat{\chi})$. Будем предполагать, что для всех $\chi \in \mathcal{B}$ и для всех $\psi \in \mathcal{A} \cap \mathcal{B}$ выполняется неравенство

$$l_h(\psi, \chi) \geq l_h(\psi, \chi_{proj}) + l_h(\chi_{proj}, \chi).$$

Для пояснения этого требования рассмотрим частный случай, когда $\mathbb{R}^d = \mathbb{R}^3$, $\mathcal{A} \subset \mathbb{R}^2$, $\mathcal{B} \subset \mathbb{R}^3$, $\mathcal{A} \cap \mathcal{B} \subset \mathbb{R}^2$ и l_h — квадрат метрики Евклида. Точки ψ , χ , χ_{proj} образуют треугольник. Обозначим θ угол при вершине χ_{proj} и запишем теорему косинусов

$$l_h(\psi, \chi) = l_h(\psi, \chi_{proj}) + l_h(\chi_{proj}, \chi) - 2\sqrt{l_h(\psi, \chi_{proj})}\sqrt{l_h(\chi_{proj}, \chi)} \cos \theta.$$

Поскольку χ_{proj} является проекцией, то угол θ не может быть острым. Он прямой, если проецируемая точка χ находится «над» множеством $\mathcal{A} \cap \mathcal{B}$, и тупой, если

точка находится «в стороне». Таким образом получаем, что $\cos \theta \leq 0$, а значит, последнее слагаемое в теореме косинусов неотрицательное. Исключая его и заменяя знак равенства на знак нестрогого неравенства, получаем, что предположение 4 соответствует естественным свойствам проекции.

Введя предположения 1–4, сформулируем теорему.

Теорема 2 Пусть выполнены предположения 1–4. Тогда пара стратегий $(\chi_{proj}, \chi_{proj})$ является равновесием Нэша в игре (6) и седловой точкой функции $L(\hat{\varphi}, \chi_{T+1}) = l_h(\chi_{T+1}, \hat{\varphi}) - l_h(\chi_{T+1}, \hat{\chi})$. Цена игры при этом равна $V = -l_h(\chi_{proj}, \hat{\chi})$.

Доказательство. Найдем седловую точку функции $L(\hat{\varphi}, \chi_{T+1})$ в соответствии с определением 4 и теоремой 1. Найдем максимум этой функции по второму аргументу при $\hat{\varphi} = \chi_{proj}$. Для этого воспользуемся предположением 4:

$$l_h(\psi, \chi) \geq l_h(\psi, \chi_{proj}) + l_h(\chi_{proj}, \chi).$$

Применяя неравенство к функции потерь L игрока 1 (подставляем $\psi = \chi_{T+1}$, $\chi = \hat{\chi}$), получаем

$$\begin{aligned} L(\chi_{proj}, \chi_{T+1}) &= l_h(\chi_{T+1}, \chi_{proj}) - l_h(\chi_{T+1}, \hat{\chi}) \leq l_h(\chi_{T+1}, \hat{\chi}) - l_h(\chi_{proj}, \hat{\chi}) - l_h(\chi_{T+1}, \hat{\chi}) = \\ &= -l_h(\chi_{proj}, \hat{\chi}) \quad \text{для всех } \chi_{T+1} \in \mathcal{A} \cap \mathcal{B}. \end{aligned}$$

Заметим также, что из предположения 2 вытекает

$$L(\chi_{proj}, \chi_{proj}) = l_h(\chi_{proj}, \chi_{proj}) - l_h(\chi_{proj}, \hat{\chi}) = -l_h(\chi_{proj}, \hat{\chi}).$$

Приходим к выводу, что

$$L(\chi_{proj}, \chi_{T+1}) \leq L(\chi_{proj}, \chi_{proj}) \quad \text{для всех } \chi_{T+1} \in \mathcal{A} \cap \mathcal{B}.$$

Следовательно, максимум по второму аргументу достигается при $\chi_{T+1} = \chi_{proj}$.

Минимум по первому аргументу при $\chi_{T+1} = \chi_{proj}$ находим, используя предположение 2, из соотношения

$$\arg \min_{\hat{\varphi} \in \mathcal{A} \cap \mathcal{B}} L(\hat{\varphi}, \chi_{proj}) = \arg \min_{\hat{\varphi} \in \mathcal{A} \cap \mathcal{B}} l_h(\chi_{proj}, \hat{\varphi}) - l_h(\chi_{proj}, \hat{\chi}).$$

Второе слагаемое не зависит от $\hat{\varphi}$, а по предположению 2 функция суммарных потерь неотрицательна и обращается в ноль только при равенстве аргументов, поэтому получаем

$$\arg \min_{\hat{\varphi} \in \mathcal{A} \cap \mathcal{B}} L(\hat{\varphi}, \chi_{proj}) = \chi_{proj}.$$

Таким образом, получаем, что

$$L(\chi_{proj}, \chi_{T+1}) \leq L(\chi_{proj}, \chi_{proj}) \leq L(\hat{\varphi}, \chi_{proj}), \quad \chi_{T+1}, \hat{\varphi} \in \mathcal{A} \cap \mathcal{B}.$$

Следовательно, $(\chi_{proj}, \chi_{proj})$ — седловая точка функции L . И эта пара является равновесием Нэша в игре (6) и цена игры выражается как

$$V = \min_{\hat{\varphi} \in \mathcal{A} \cap \mathcal{B}} \max_{\chi_{T+1} \in \mathcal{A}} L(\hat{\varphi}, \chi_{T+1}) = \max_{\chi_{T+1} \in \mathcal{A}} \min_{\hat{\varphi} \in \mathcal{A} \cap \mathcal{B}} L(\hat{\varphi}, \chi_{T+1}) = L(\chi_{proj}, \chi_{proj}) = -l_h(\chi_{proj}, \hat{\chi}).$$

■

Следствие 1 *Использование в качестве вектора согласованных прогнозов $\hat{\varphi}$ проекции вектора независимых прогнозов $\hat{\chi} \in \mathbb{R}^d$ на множество $\mathcal{A} \cap \mathcal{B}$, инициализированной функцией суммарных потерь l_h , гарантирует значение функции суммарных потерь не большее, чем при использовании вектора независимых прогнозов $\hat{\chi}$.*

Доказательство. По теореме 2 цена игры (6) равна значению функции потерь игрока 1 в точке $(\chi_{proj}, \chi_{proj})$ и неположительна в силу предположения 2

$$V = L(\chi_{proj}, \chi_{proj}) = -l_h(\chi_{proj}, \hat{\chi}) \leq 0.$$

А выбор рассматриваемой функции потерь игрока 1 $L(\hat{\varphi}, \chi_{T+1}) = l_h(\chi_{T+1}, \hat{\varphi}) - l_h(\chi_{T+1}, \hat{\chi})$ был обусловлен тем, что ее знак совпадает со знаком изменения суммарных потерь при переходе от вектора независимых прогнозов $\hat{\chi}$ к вектору согласованных прогнозов $\hat{\varphi}$. Следовательно, при $\chi_{T+1} = \chi_{proj}$ суммарные потери при согласованных прогнозах меньше, чем при независимых.

Согласно определению 4 равновесия Нэша

$$L(\chi_{proj}, \chi_{T+1}) \leq L(\chi_{proj}, \chi_{proj}) \leq 0 \quad \text{для любых } \chi_{T+1} \in \mathcal{A} \cap \mathcal{B}.$$

Поэтому при любом векторе действительных значений $\chi_{T+1} \in \mathcal{A} \cap \mathcal{B}$ согласованные прогнозы $\hat{\varphi}$ оказываются предпочтительнее независимых прогнозов $\hat{\chi}$. ■

3.3 Алгоритм согласования прогнозов GТОр

Согласно следствию 1 оптимальным выбором вектора согласованных прогнозов $\hat{\varphi}$ является проекция вектора независимых прогнозов $\hat{\chi}$ на множество $\mathcal{A} \cap \mathcal{B}$, инициированная функцией суммарных потерь l_h . Множество $\mathcal{A} \cap \mathcal{B}$ содержит векторы размерности d , удовлетворяющие структуре иерархии, так как множество \mathcal{A} задается ограничениями-равенствами, порожденными матрицей связи иерархии S . В то же время $\mathcal{A} \cap \mathcal{B}$ содержит d -мерные векторы, удовлетворяющие ограничениям-неравенствам из множества \mathcal{B} . Таким образом, задача поиска проекции — это оптимизационная задача с ограничениями типа равенства и неравенства:

$$\begin{cases} l_h(\chi, \hat{\chi}) \rightarrow \min_{\chi}, \\ \chi \in \mathcal{A}, & \text{ограничения-равенства;} \\ \chi \in \mathcal{B}, & \text{ограничения-неравенства.} \end{cases} \quad (7)$$

Алгоритм согласования прогнозов иерархических временных рядов GТОр заключается в решении оптимизационной задачи (7). Достоинства этого алгоритма заключаются в том, что он требует от вектора независимых прогнозов $\hat{\chi}$ лишь принадлежности множеству \mathcal{B} , и не требует несмещенности независимых прогнозов, а следовательно, для получения независимых прогнозов можно использовать любой алгоритм прогнозирования. Также GТОр не требует оценки погрешностей независимых прогнозов. Самое важное, что GТОр обеспечивает неухудшение качества прогнозирования при замене независимых прогнозов на согласованные прогнозы. При этом на структуру иерархии, ограничения-неравенства на прогнозы и функцию суммарных потерь накладываются лишь общие ограничения, гарантирующие существование решения оптимизационной задачи (7). Еще одно достоинство алгоритма GТОр в том, что он позволяет согласовывать прогнозы для иерархий любой сложности одновременно по всем уровням, учитывая все связи в иерархии и решая одну оптимизационную задачу. От сложности иерархии и количества временных рядов и уровней в оптимизационной задаче зависит число переменных и ограничений.

3.4 Функции потерь, применимые для GTOp

В этом разделе будет описано семейство функций двух переменных, удовлетворяющих предположениям 1–4, и приведен ряд примеров функций из этого семейства, которые можно использовать в задаче согласования прогнозов в качестве функции суммарных потерь l_h (5). Все эти функции двух переменных называются дивергенциями Брегмана [9, 59] и порождаются функциями одной переменной, обладающими следующими свойствами.

Определение 7 *Функцией Лежандра [9] называется функция $F: \mathcal{B} \rightarrow \mathbb{R}$, которая удовлетворяет следующим условиям:*

- $\mathcal{B} \subseteq \mathbb{R}^d$ — непустое множество, и внутренность \mathcal{B} выпукла;
- F — строго выпуклая функция с непрерывной первой производной на множестве \mathcal{B} ;
- если $\chi_1, \chi_2, \dots \in \mathcal{B}$ — последовательность, сходящаяся к граничной точке \mathcal{B} , то $\|\nabla F(\chi_n)\| \rightarrow \infty$ при $n \rightarrow \infty$.

Определение 8 *Дивергенцией Брегмана, порожденной функцией Лежандра $F: \mathcal{B} \rightarrow \mathbb{R}$, называется неотрицательная функция $D_F: \mathcal{B} \times \text{int}(\mathcal{B}) \rightarrow \mathbb{R}$, определенная как*

$$D_F(\mathbf{u}, \mathbf{v}) = F(\mathbf{u}) - F(\mathbf{v}) - (\mathbf{u} - \mathbf{v})\nabla F(\mathbf{v}).$$

Свойства дивергенции Брегмана:

- для всех \mathbf{u}, \mathbf{v} выполнено $D_F(\mathbf{u}, \mathbf{v}) \geq 0$. Это следует из выпуклости функции F ;
- $D_F(\mathbf{u}, \mathbf{v})$ выпукла по первому аргументу \mathbf{u} , но не обязательно выпукла по второму аргументу \mathbf{v} ;
- для любых $\alpha, \beta \in \mathbb{R}$ и любых функций Лежандра F_1, F_2 выполнено $D_{\alpha F_1 + \beta F_2}(\mathbf{u}, \mathbf{v}) = \alpha D_{F_1}(\mathbf{u}, \mathbf{v}) + \beta D_{F_2}(\mathbf{u}, \mathbf{v})$.

Определение 9 *Пусть $F: \mathcal{B} \rightarrow \mathbb{R}$ — функция Лежандра и $\mathcal{A} \subset \mathbb{R}^d$ — замкнутое выпуклое множество, такое что $\mathcal{A} \cap \mathcal{B} \neq \emptyset$. Проекция Брегмана \mathbf{w}' точки $\mathbf{w} \in \text{int}(\mathcal{B})$ на множество \mathcal{A} — это*

$$\mathbf{w}' = \arg \min_{\mathbf{u} \in \mathcal{A} \cap \mathcal{B}} D_F(\mathbf{u}, \mathbf{w}).$$

Теорема 3 Для всех функций Лежандра $F: \mathcal{B} \rightarrow \mathbb{R}$, для всех замкнутых выпуклых множеств $\mathcal{A} \subset \mathbb{R}^d$, имеющих непустое пересечение $\mathcal{A} \cap \mathcal{B} \neq \emptyset$, и для всех точек $\mathbf{w} \in \text{int}(\mathcal{B})$ проекция Брегмана точки \mathbf{w} на множество \mathcal{A} существует и единственна.

Доказательство этой теоремы приведено в статье [59].

Теорема 4 Пусть F — функция Лежандра. Для всех $\mathbf{w} \in \text{int}(\mathcal{B})$ и для всех замкнутых выпуклых множеств $\mathcal{A} \subseteq \mathbb{R}^d$ с непустым пересечением $\mathcal{A} \cap \mathcal{B} \neq \emptyset$, если \mathbf{w}' — проекция Брегмана точки \mathbf{w} на множество \mathcal{A} ($\mathbf{w}' = \arg \min_{\mathbf{v} \in \mathcal{A} \cap \mathcal{B}} D_F(\mathbf{v}, \mathbf{w})$), то верно неравенство

$$D_F(\mathbf{u}, \mathbf{w}) \geq D_F(\mathbf{u}, \mathbf{w}') + D_F(\mathbf{w}', \mathbf{w}) \quad \text{для всех } \mathbf{u} \in \mathcal{A}.$$

Доказательство этого факта можно найти в [9].

Соотнесем перечисленные свойства дивергенции Брегмана и предположения 1–4. Определение 8 и свойство 1 дивергенции Брегмана обеспечивают выполнение предположения 2 о знаке функции суммарных потерь. В определении 9 и теореме 3 предполагается выпуклость и замкнутость множеств \mathcal{A} и \mathcal{B} и их непустое пересечение, как и в предположении 1. Теорема 4 гарантирует выполнение предположения 3 о выпуклости множеств и существовании и единственности проекции. Наконец, теорема 4 гарантирует выполнение предположения 4. Следовательно, для функций суммарных потерь l_h (5), являющихся дивергенциями Брегмана, выполнены все условия теоремы 2. Использование в качестве вектора согласованных прогнозов $\hat{\varphi}$ проекции Брегмана вектора независимых прогнозов $\hat{\chi} \in \mathcal{B}$, где множество \mathcal{B} определено по формуле (4) и является выпуклым и замкнутым, имеющим непустое пересечение с множеством \mathcal{A} (3), на множество \mathcal{A} , включающее векторы прогнозов, удовлетворяющих структуре иерархии, гарантирует неухудшение качества прогнозов.

Следующие функции являются дивергенциями Брегмана и могут быть использованы в качестве функций суммарных потерь (5) при согласовании прогнозов.

Квадрат евклидова расстояния $D_F(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|^2$ — канонический пример дивергенции Брегмана, порождается функцией $F(\mathbf{u}) = \|\mathbf{u}\|^2$.

Квадрат расстояния Махаланобиуса $D_F(\mathbf{u}, \mathbf{v}) = \frac{1}{2}(\mathbf{u} - \mathbf{v})^\top Q(\mathbf{u} - \mathbf{v})$ — обобщение евклидова расстояния, порождается квадратичной формой $F(\mathbf{u}) = \frac{1}{2}\mathbf{u}^\top Q\mathbf{u}$.

Обобщенная дивергенция Кульбака–Лейблера $D_F(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^n u_i \log \frac{u_i}{v_i} - \sum_{i=1}^n u_i + \sum_{i=1}^n v_i$

порождается функцией $F(\mathbf{u}) = \sum_{i=1}^n u_i \log u_i - \sum_{i=1}^n u_i$.

Расстояние Itakura–Saito $D_F(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^n \left(\frac{u_i}{v_i} - \log \frac{u_i}{v_i} - 1 \right)$ порождается функцией

$$F(\mathbf{u}) = - \sum_{i=1}^n \log u_i.$$

Homogeneous loss function $D_F(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^n |u_i|^a - \sum_{i=1}^n |v_i|^a - a \sum_{i=1}^n \text{sign}(v_i) |v_i|^{a-1} (u_i - v_i)$

порождается функцией $F(\mathbf{u}) = \sum_{i=1}^n |u_i|^a$, $a > 1$, впервые предложена в [60].

Недифференцируемость порождающей функции в нуле можно не принимать во внимание, если прогнозируемые величины обладают непрерывным распределением. При $a = 2$ эта функция совпадает с квадратом евклидова расстояния.

Non-homogeneous loss function $D_F(\mathbf{u}, \mathbf{v}) = \frac{2}{a^2} \sum_{i=1}^n (e^{au_i} - e^{av_i}) - \frac{2}{a} \sum_{i=1}^n e^{av_i} (u_i - v_i)$

порождается функцией $F(\mathbf{u}) = \frac{2}{a^2} \sum_{i=1}^n e^{au_i}$, $a \neq 0$, впервые предложена в [61]. При $a \rightarrow 0$ значения этой функции стремятся к значениям квадрата евклидова расстояния.

3.5 Модификация GТОр с учетом надежности независимых прогнозов — MGТОр

Алгоритм согласования прогнозов GТОр сводит решение задачи согласования прогнозов иерархических временных рядов к решению оптимизационной задачи (7) с ограничениями типа равенства и неравенства. В ходе решения задачи (7) в качестве вектора согласованных прогнозов $\hat{\varphi}$ находится вектор, который максимально близок к вектору независимых прогнозов $\hat{\chi}$ и при этом удовлетворяет структуре иерархии и физическим ограничениям. Близость векторов $\hat{\varphi}$ и $\hat{\chi}$ оценивается в терминах функции потерь прогнозирования l_h (5). При этом функция потерь $l_h(\chi_{T+1}, \hat{\chi})$ является дивергенцией Брегмана и учитывает стоимость ошибок прогнозирования для каждого временного ряда, поскольку веса u слагаемых, соответствующих разным временным рядам, могут быть различными. При этом в процессе согласования прогнозы рядов с наиболее высокой стоимостью ошибки будут корректироваться минимально, а прогнозы рядов с меньшей стоимостью ошибки прогнози-

рования будут подвергаться коррекции в большей степени. Однако на практике временные ряды с высокой стоимостью ошибки прогнозирования не всегда совпадают с теми рядами, для которых полученные независимые прогнозы наиболее точны. Если же в ходе процедуры согласования в наименьшей степени корректировать наиболее точные прогнозы, то это может обеспечить более высокое качество прогнозирования по всей иерархии, а именно, более существенное уменьшение значения функции потерь после согласования $l_h(\chi_{T+1}, \hat{\varphi})$ по сравнению с потерями от независимых прогнозов $l_h(\chi_{T+1}, \hat{\chi})$.

Ввиду вышесказанного введем функцию потерь, которая будет использоваться для проведения процедуры согласования прогнозов алгоритмом GTOp и которая может не совпадать с функцией потерь прогнозирования l_h (5)

$$l_r(\chi, \hat{\chi}). \quad (8)$$

Тогда в ходе согласования прогнозов алгоритмом GTOp будет решаться оптимизационная задача

$$\begin{cases} l_r(\chi, \hat{\chi}) \rightarrow \min_{\chi}, \\ \chi \in \mathcal{A}, & \text{ограничения-равенства;} \\ \chi \in \mathcal{B}, & \text{ограничения-неравенства.} \end{cases} \quad (9)$$

Если функция потерь согласования l_r (8) является дивергенцией Брегмана, то по теореме 2 и следствию 1 после нахождения вектора согласованных прогнозов $\hat{\varphi}$ путем решения оптимизационной задачи (9) для нее выполняется неравенство

$$l_r(\chi_{T+1}, \hat{\varphi}) \leq l_r(\chi_{T+1}, \hat{\chi}).$$

Функция l_r , например, может являться взвешенным евклидовым расстоянием с весами, выбранными обратно пропорциональными ширине доверительных интервалов прогнозов для каждого временного ряда, как это предложено в [4]. Также веса для временных рядов в функции l_r (8) можно настраивать с помощью кросс-валидации, добиваясь выполнения неравенства

$$l_h(\chi_{T+1}, \hat{\varphi}) \leq l_h(\chi_{T+1}, \hat{\chi})$$

для функции потерь прогнозирования l_h (5) как можно с большим запасом. Функция потерь прогнозирования l_h (5) при этом может и не являться дивергенцией Брегмана.

4 Вычислительный эксперимент

В экспериментальной части демонстрируется работа алгоритмов согласования прогнозов GTOp и MGTOp на двух наборах данных. Первый набор описывает посуточные объемы отгрузки набора грузов в заданных узлах РЖД. Эти данные предоставлены ОАО РЖД. Второй набор данных описывает почасовые объемы потребления электроэнергии в ряде регионов Канады и предоставлен организаторами конкурса, проведенного в 2012 году на Kaggle [62]. Независимые прогнозы временных рядов обоих наборов данных получены с помощью модели *ARIMA* с подобранными параметрами для каждого набора данных.

В первой части раздела приводится подробное описание наборов данных и их иерархических структур, во второй части демонстрируется, что согласованные прогнозы, полученные алгоритмом GTOp с квадратом евклидового расстояния в качестве функции потерь прогнозирования l_h (5) обладают всеми свойствами, которые были обнаружены в теоретических выкладках. В третьей части в качестве функции потерь прогнозирования l_h (5) используется квадрат евклидового расстояния, а в качестве функции потерь согласования l_r (8) используется квадрат взвешенного евклидового расстояния для получения согласованных прогнозов алгоритмом MGTOp. Веса для функции l_r подбираются с помощью кросс-валидации. А также демонстрируется качество согласованных прогнозов, полученных с помощью алгоритма MGTOp, где веса в функции потерь согласования l_r выбраны обратно пропорциональными погрешностям независимых прогнозов.

4.1 Экспериментальные данные

4.1.1 Посуточная загруженность узлов РЖД

В эксперименте использованы данные о посуточной загруженности железнодорожных узлов РЖД с 1 января 2007 года по 22 апреля 2008 года. В табл. 1 приведен пример записи базы данных.

Коды станций представляют собой шестизначные числа. Станции, в коде которых две первые цифры совпадают, входят в одну железнодорожную ветку. Станций отправления — 1566, станций назначения — 1902, веток — 98. Код груза — натураль-

Таблица 1: Вид записи базы данных железнодорожных перевозок

Дата погрузки	Станция отправления	Станция назначения	Количество вагонов	Код груза	Род вагона	Суммарный вес груза	Признак маршрутной отправки
2007-01-01	020108	932902	1	1	216	56	9

ное число от 1 до 37; также имеются перевозки, где код груза не указан. Род вагона — натуральное число, в имеющихся данных 75 различных родов вагонов.

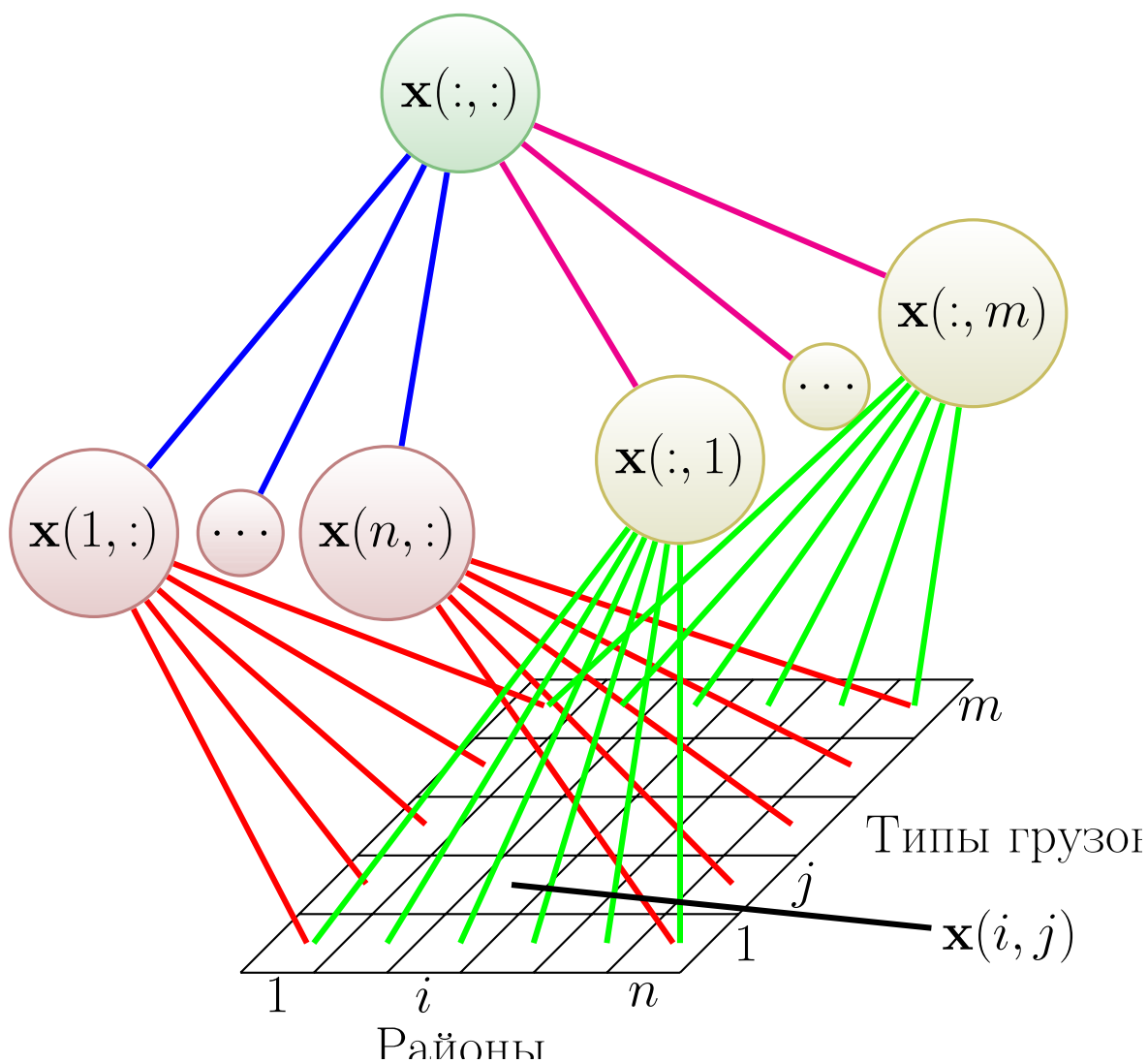


Рис. 3: Иерархическая структура для данных РЖД

Из имеющихся данных были сформированы временные ряды, описывающие отправление 37 различных типов груза со станций 98 железнодорожных веток посуточно. Экспериментальные данные удовлетворяют структуре, изображенной на рис. 3. Как видно из рисунка, иерархия не является плоской и содержит три уровня временных рядов. Временные ряды нижнего уровня этой иерархии имеют два индекса, соответствующих номеру ветки и коду груза: $\mathbf{x}(i, j)$, $i = 1, \dots, n$, $j = 1, \dots, m$, где $n = 98$ — число веток, а $m = 37$ — количество грузов. На среднем уровне — два семейства временных рядов. Временные ряды, соответствующие суммарному отправлению всех грузов с каждой ветки, обозначаются $\mathbf{x}(i, :)$, $i = 1, \dots, n$. Ряды среднего уровня, соответствующие суммарному отправлению со всех веток каждого из грузов, обозначаются $\mathbf{x}(:, j)$, $j = 1, \dots, m$. Временной ряд верхнего уровня обозначается $\mathbf{x}(:, :)$. Условие согласованности для трехуровневой иерархии задается равенствами

$$\begin{aligned}
x_t(:, :) &= \sum_{i=1}^n x_t(i, :), \quad t = 1, \dots, T; \\
x_t(:, :) &= \sum_{j=1}^m x_t(:, j), \quad t = 1, \dots, T; \\
x_t(i, :) &= \sum_{j=1}^m x_t(i, j), \quad t = 1, \dots, T, \quad i = 1, \dots, n; \\
x_t(:, j) &= \sum_{i=1}^n x_t(i, j), \quad t = 1, \dots, T, \quad j = 1, \dots, m.
\end{aligned} \tag{10}$$

Векторная запись срезов иерархии, независимых прогнозов и согласованных прогно-

зов имеет размерность $d = 1 + n + m + nm$ и выглядит следующим образом:

$$\mathbf{x}_t = \begin{pmatrix} x_t(:, :) \\ x_t(1, :) \\ \dots \\ x_t(n, :) \\ x_t(:, 1) \\ \dots \\ x_t(:, m) \\ x_t(1, 1) \\ \dots \\ x_t(1, m) \\ \dots \\ x_t(n, 1) \\ \dots \\ x_t(n, m) \end{pmatrix}, \quad \hat{\mathbf{x}} = \begin{pmatrix} \hat{x}(:, :) \\ \hat{x}(1, :) \\ \dots \\ \hat{x}(n, :) \\ \hat{x}(:, 1) \\ \dots \\ \hat{x}(:, m) \\ \hat{x}(1, 1) \\ \dots \\ \hat{x}(1, m) \\ \dots \\ \hat{x}(n, 1) \\ \dots \\ \hat{x}(n, m) \end{pmatrix}, \quad \hat{\varphi} = \begin{pmatrix} \hat{y}(:, :) \\ \hat{y}(1, :) \\ \dots \\ \hat{y}(n, :) \\ \hat{y}(:, 1) \\ \dots \\ \hat{y}(:, m) \\ \hat{y}(1, 1) \\ \dots \\ \hat{y}(1, m) \\ \dots \\ \hat{y}(n, 1) \\ \dots \\ \hat{y}(n, m) \end{pmatrix}.$$

Матрица связей для условия (10) имеет размер $(2 + n + m) \times (1 + n + m + nm)$.

$$S = \left(\begin{array}{c|ccc|ccc|ccc|ccc|ccc|ccc} -1 & 1 & \dots & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ -1 & 0 & \dots & 0 & 1 & \dots & 1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ \hline 0 & -1 & \dots & 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ \dots & & \dots & & & \dots & & & \dots & & & & \dots & & & \dots & & \dots & & \dots \\ 0 & 0 & \dots & -1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 1 & 1 & \dots & 1 \\ \hline 0 & 0 & \dots & 0 & -1 & \dots & 0 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & \dots & 1 & 0 & \dots & 0 \\ \dots & & \dots & & & \dots & & & \dots & & & & \dots & & & \dots & & \dots & & \dots \\ 0 & 0 & \dots & 0 & 0 & \dots & -1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 & \dots & 0 & 0 & \dots & 1 \end{array} \right).$$

Требуется, чтобы все прогнозы были неотрицательны, поэтому множество \mathcal{B} (4) задается как

$$\mathcal{B} = \{\mathbf{x} \in \mathbb{R}^d \mid \mathbf{x}(i) \in [0, +\infty], i = 1, \dots, d\}.$$

4.1.2 Почасовое потребление электроэнергии

Из всех данных, предоставленных организаторами конкурса [62], было использовано почасовое потребление электроэнергии с 00:00 29 ноября 2007 г. по 05:00 30 июня 2008 г.

Данные описывают объемы потребления электроэнергии в 20 регионах Канады и объем суммарного потребления за этот же период. Иерархическая структура, соответствующая этим временным рядам изображена на рис. 4. Эта иерархия плоская и содержит два уровня временных рядов.

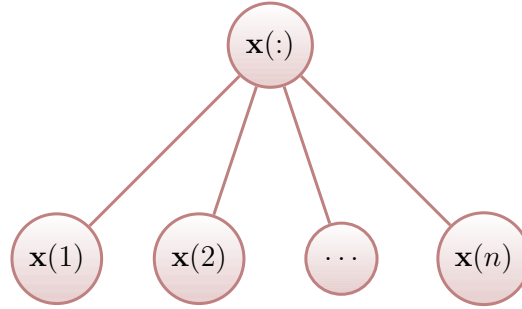


Рис. 4: Иерархическая структура для данных по потреблению электроэнергии

На нижнем уровне иерархии $n = 20$ временных рядов $\mathbf{x}(i)$, $i = 1, \dots, n$, соответствующих потреблению в регионах, на верхнем уровне один временной ряд $\mathbf{x}(:)$, описывающий суммарное потребление электроэнергии. Условие согласованности для этого набора данных записывается как

$$x_t(:) = \sum_{i=1}^n x_t(i), \quad t = 1, \dots, T. \quad (11)$$

Векторная запись срезов иерархии, независимых прогнозов и согласованных прогнозов имеет размерность $d = 1 + n$ и выглядит следующим образом:

$$\boldsymbol{\chi}_t = \begin{pmatrix} x_t(:) \\ x_t(1) \\ \dots \\ x_t(n) \end{pmatrix}, \quad \hat{\boldsymbol{\chi}} = \begin{pmatrix} \hat{x}(:) \\ \hat{x}(1) \\ \dots \\ \hat{x}(n) \end{pmatrix}, \quad \hat{\boldsymbol{\varphi}} = \begin{pmatrix} \hat{y}(:) \\ \hat{y}(1) \\ \dots \\ \hat{y}(n) \end{pmatrix}.$$

Матрица связей для условия (11) имеет размер $1 \times (n + 1)$

$$S = \begin{pmatrix} -1 & 1 & \dots & 1 \end{pmatrix}.$$

Требуется, чтобы все прогнозы были неотрицательны, поэтому множество \mathcal{B} (4) задается как

$$\mathcal{B} = \{\boldsymbol{\chi} \in \mathbb{R}^d \mid \chi(i) \in [0, +\infty], i = 1, \dots, d\}.$$

4.2 Демонстрация свойств алгоритма GTOp

4.2.1 Оптимизационная задача

В качестве функции потерь l_h (5) для алгоритма согласования GTOp был использован квадрат евклидового расстояния. Для нахождения вектора согласованных прогнозов $\hat{\varphi}$ решалась оптимизационная задача

$$\begin{cases} \|\chi - \hat{\chi}\|_2^2 \rightarrow \min_{\chi}, \\ S\chi = 0, \\ \chi \geq 0. \end{cases}$$

4.2.2 Результаты

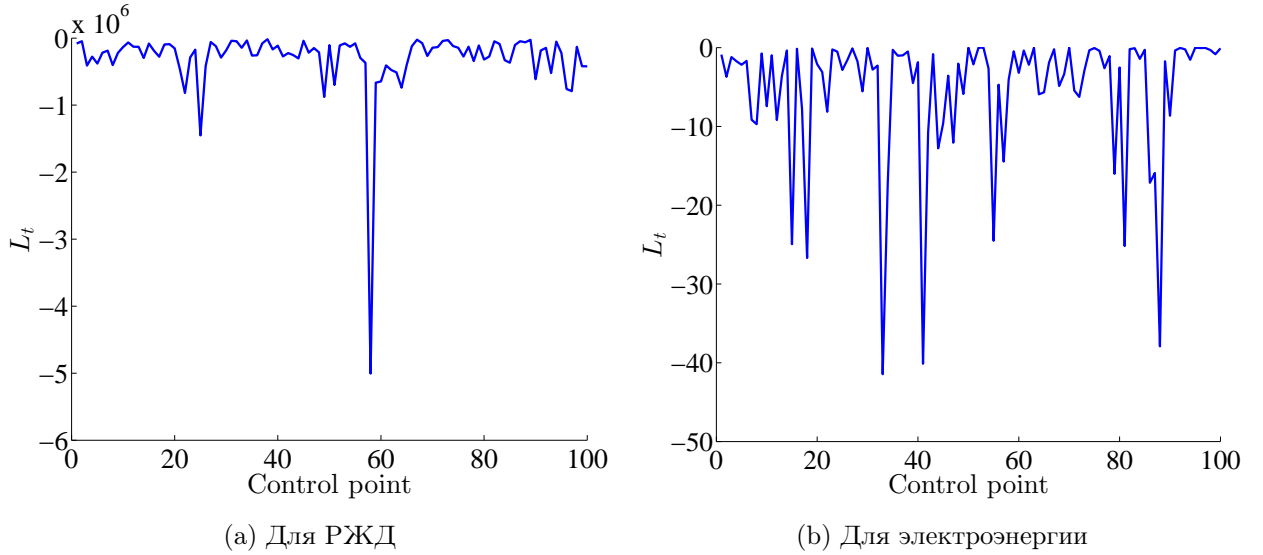


Рис. 5: Изменение суммарных потерь L_t

Независимые прогнозы для всех временных рядов были получены с помощью модели *ARIMA*. Процедура согласования проводилась в 100 контрольных точках. Для каждого отсчета времени строился вектор независимых прогнозов $\hat{\chi}$ и вектор согласованных прогнозов $\hat{\varphi}$. Для каждого вектора вычислялось значение функции суммарных потерь, затем вычислялись потери игрока 1 в игре (6), равные разности суммарных потерь при использовании согласованных прогнозов и суммарных потерь при использовании независимых прогнозов:

$$L_t = l_h(\chi_t, \hat{\varphi}) - l_h(\chi_t, \hat{\chi}), \quad t = T - 100 + 1, \dots, T.$$

Значения этой величины изображены на рис. 5. По оси абсцисс во всех отложены номера контрольных точек. По оси ординат отложена разности суммарных потерь при использовании согласованных прогнозов и суммарных потерь при использовании независимых прогнозов L_t . Теоретические выкладки подтверждаются на практике. Для плоской двухуровневой иерархии есть случаи, когда суммарные потери при переходе к согласованным прогнозам не изменяются, и случаи, когда суммарные потери уменьшаются. Для неплоской трехуровневой иерархии суммарные потери во всех контрольных точках уменьшаются.

4.3 Выбор весов функции потерь для MGTOp

В этой части раздела вектор проводится процедура согласования прогнозов алгоритмом MGTOp и качество полученных прогнозов сравнивается с прогнозами, полученными при помощи алгоритма восходящего согласования [5] и оптимальной регрессии [2]. Алгоритм восходящего согласования был применен в стандартной реализации. Алгоритм оптимальной регрессии в исходном описании не позволяет учитывать физические ограничения на прогнозы, поэтому в решаемую в ходе выполнения этого алгоритма задачу линейной регрессии были дополнительно введены ограничения-неравенства, чтобы обеспечить выполнение физических ограничений.

4.3.1 Оптимизационная задача

В этой части эксперимента в качестве функции потерь прогнозирования l_h (5) был по-прежнему использован квадрат евклидового расстояния, а в качестве функции потерь согласования l_r (8) был использован квадрат взвешенного евклидового расстояния (частный случай квадрата расстояния Махаланобиса). Веса в функции l_r для временных рядов промежуточного и верхнего уровня были выбраны единичными и зафиксированы, а веса для временных рядов нижнего уровня были выбраны равными между собой и подбирались с помощью кросс-валидации. Также демонстрируется работа алгоритма, когда веса в функции l_r выбираются обратно пропорциональными погрешностям независимых прогнозов для каждого временного ряда. Для нахождения вектора согласованных прогнозов алгоритмом MGTOp решалась

$$\begin{cases} l_r(\chi - \hat{\chi}) \rightarrow \min, \\ S\chi = 0, \\ \chi \geq 0. \end{cases}$$

4.3.2 Результаты

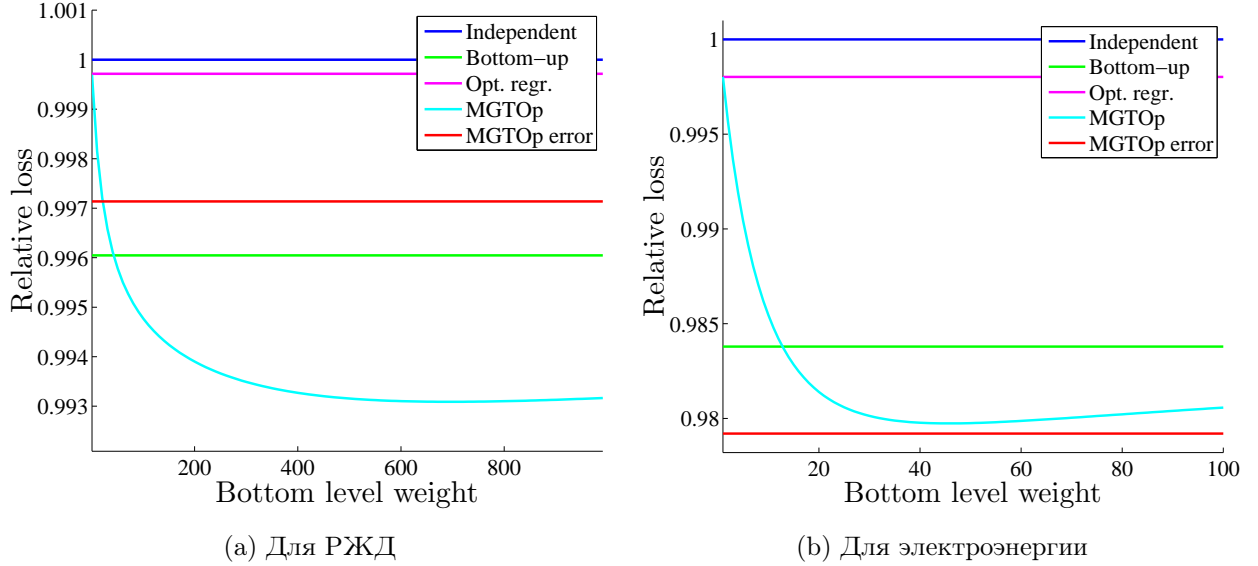


Рис. 6: Выбор весов функции потерь для MGTOp

Независимые прогнозы для всех временных рядов были получены с помощью модели *ARIMA*. Процедура согласования проводилась в 100 контрольных точках. Для каждого отсчета времени строился вектор независимых прогнозов $\hat{\chi}$ и вектор согласованных прогнозов $\hat{\varphi}$ при помощи трех алгоритмов согласования. Вектор согласованных прогнозов для алгоритма MGTOp вычислялся многократно для различных значений весов в функции потерь согласования l_r с целью выбора наиболее подходящего значения весов.

Для независимых прогнозов и согласованных прогнозов, полученных разными алгоритмами, вычислялись средние потери по всем контрольным точкам. На рис. 6 построены графики средней относительной ошибки прогнозов. По оси абсцисс отложен вес для временных рядов нижнего уровня иерархии в функции потерь согласования l_r (8). По оси ординат отложено среднее по всем контрольным точкам значение функции потерь прогнозирования l_h (5) для всех векторов прогнозов, поделенное на

среднее значение функции l_h для независимых прогнозов:

$$\text{Relative loss}(\text{algorithm}) = \frac{\sum_{t=T-k+1}^T l_h(\mathbf{x}_t, \hat{\varphi}_{\text{algorithm}})}{\sum_{t=T-k+1}^T l_h(\mathbf{x}_t, \hat{\mathbf{x}})}.$$

Поскольку алгоритмы согласования, кроме MGTOp, не предоставляют возможности менять веса у временных рядов нижнего уровня, то для них графики относительной ошибки горизонтальны. Относительная ошибка на графике для алгоритма MGTOp при единичном весе соответствует относительной ошибке алгоритма GTOp, у которого $l_r = l_h$. Для ошибки согласованных прогнозов, полученных алгоритмом MGTOp, на графиках виден хорошо выраженный минимум. Из этого следует, что при помощи выбора весов для функции l_r можно добиться качества прогнозов, превосходящего все качество прогнозов всех остальных алгоритмов согласования и качества независимых прогнозов. На обоих графиках красным построен график относительных потерь для согласованных прогнозов, полученных алгоритмом MGTOp с функцией потерь согласования l_r , веса у которой выбраны обратно пропорциональными среднеквадратичной погрешности независимых прогнозов для каждого временного ряда.

Таблица 2: Средние потери прогнозирования объемов отгрузки в узлах РЖД, $\times 10^8$

Уровень иерархии	Независимые прогнозы	Восходящее согласование	Оптимальная регрессия	MGTOp (веса 700)
Вся иерархия	10.038	9.999	10.035	9.969
Верхний уровень	2.858	2.868	2.856	2.840
Средний уровень, ветки	2.549	2.486	2.545	2.487
Средний уровень, грузы	2.338	2.351	2.340	2.348
Нижний уровень	2.294	2.294	2.294	2.294

Для рядов нижнего уровня с весом 700 в табл. 2 и для весов, обратно пропорциональным среднеквадратичной погрешности независимых прогнозов, в табл. 3 представлены средние значения функции потерь прогнозирования l_h (квадрата евклидова расстояния) для иерархии в целом и для подмножеств временных рядов

Таблица 3: Средние потери прогнозирования потребления электроэнергии

Уровень иерархии	Независимые прогнозы	Восходящее согласование	Оптимальная регрессия	MGTOp
Вся иерархия	2727	2683	2722	2670
Верхний уровень	2083	2039	2076	2029
Нижний уровень	644	644	646	642

каждого уровня иерархии. Для подмножеств временных рядов функция потерь вычислялась как квадрат евклидова расстояния между векторами соответствующей размерности. Потери по иерархии в целом равны сумме потерь по всем уровням иерархии.

В табл. 2 и 3 видно, что согласованные прогнозы, полученные при помощи алгоритма MGTOp, обеспечивают наилучшее качество в целом по иерархии и на верхнем уровне, при этом не ухудшая качество прогнозирования на нижнем уровне иерархии.

Заключение

Предложен алгоритм согласования прогнозов иерархических временных рядов GТОр, который находит вектор согласованных прогнозов, удовлетворяющий условию согласованности и физическим ограничениям, максимально близкий к вектору независимых прогнозов в терминах выбранной функции потерь. Алгоритм гарантирует, что согласованные прогнозы будут обеспечивать качество прогнозирования не меньшее, чем независимые прогнозы. Также предложена модификация алгоритма, которая позволяет учитывать при проведении процедуры согласования надежность независимых прогнозов разных временных рядов, подвергая наиболее надежные прогнозы минимальной коррекции. Проведено сравнение качества согласованных прогнозов, полученных предложенным алгоритмом, с качеством согласованных прогнозов, полученных другими алгоритмами. Продемонстрировано, что разработанный алгоритм GТОр позволяет получать прогнозы более точные, чем другие алгоритмы согласования.

Список литературы

- [1] Hong Tao, Pinson Pierre, Fan Shu. Global energy forecasting competition 2012 // International Journal of Forecasting. — 2014. — Vol. 30, no. 2. — P. 357–363.
- [2] Optimal combination forecasts for hierarchical time series / Rob J. Hyndman, Roman A. Ahmed, George Athanasopoulos, Han Lin Shang // Computational Statistics & Data Analysis. — 2011. — Vol. 55, no. 9. — P. 2579–2589.
- [3] Сглаживающие алгоритмы прогнозирования / М.П. Кузнецов, А.А. Мафусалов, Н.К. Животовский и др. // Машинное обучение и анализ данных. — 2011. — Т. 1, № 1. — С. 104–112.
- [4] Стенина М.М., Стрижов В.В. Согласование агрегированных и детализированных прогнозов при решении задач непараметрического прогнозирования // Системы и средства информатики. — 2014. — Т. 24, № 2. — С. 21–34.
- [5] Shlifer E., Wolff R.W. Aggregation and proration in forecasting // Management Science. — 1979. — Vol. 25, no. 6. — P. 594–603.
- [6] Hyndman Rob J., Lee Alan J., Wang Earo. Fast computation of reconciled forecasts for hierarchical and grouped time series. — 2014. — June. — URL: <http://robjhyndman.com/papers/hgts4.pdf>.
- [7] Меньшиков И.С. Лекции по теории игр и экономическому моделированию. — М.: МЗ Пресс, 2010.
- [8] Петросян Л.А., Зенкевич Н.А., Семина Е.А. Теория игр. — Книжный Дом "Университет" М., 1998.
- [9] Cesa-Bianchi Nicolo, Lugosi Gábor et al. Prediction, learning, and games. — Cambridge University Press Cambridge, 2006. — Vol. 1.
- [10] Boyd Stephen, Vandenberghe Lieven. Convex optimization. — Cambridge university press, 2009.
- [11] Стенина М.М. Согласование прогнозов иерархических временных рядов как задача поиска равновесия Нэша в антагонистической игре // Труды 57-й научной

конференции МФТИ с международным участием, посвященной 120-летию со дня рождения П.Л. Капицы. — 2014. — Р. 75–76.

- [12] Стенина М.М. Согласование прогнозов иерархических временных рядов как задача поиска равновесия Нэша в антагонистической игре // Сборник тезисов XXII Международной научной конференции студентов, аспирантов и молодых ученых «Ломоносов-2015», секция «Вычислительная математика и кибернетика». — 2015. — Р. 72–74.
- [13] Медведникова М.М., Стрижов В.В., Кузнецов М.П. Алгоритм многоклассовой монотонной Парето-классификации с выбором признаков // Известия Тульского государственного университета. Естественные науки. — 2012. — № 3. — С. 132–141.
- [14] Кузнецов М.П., Стрижов В.В., Медведникова М.М. Алгоритм многоклассовой классификации объектов, описанных в ранговых шкалах // Научно-технический вестник С.-Пб. ПГУ. Информатика. Телекоммуникации. Управление. — 2012. — Т. 5. — С. 92–95.
- [15] Медведникова М.М., Стрижов В.В. Построение интегрального индикатора качества научных публикаций методами ко-кластеризации // Известия Тульского Государственного Университета. Естественные Науки. — 2013. — по. 1.
- [16] Газизуллина Р.К., Стенина М.М., Стрижов В.В. Прогнозирование объемов железнодорожных грузоперевозок по парам веток // Системы и средства информатики. — 2015. — Т. 25, № 1.
- [17] Стенина М.М., Стрижов В.В. Согласование прогнозов при решении задач прогнозирования иерархических временных рядов // Информатика и ее применения. — 2015. — Т. 9, № 2.
- [18] Медведникова М.М. Использование метода главных компонент при построении интегральных индикаторов // Машинное обучение и анализ данных. — 2012. — № 3. — С. 292–304.

- [19] Stenina M.M., Kuznetsov M.P., Strijov V.V. Ordinal classification using Pareto fronts // Expert Systems with Applications. — 2015. — Vol. 42, no. 14. — P. 5947 – 5953. — URL: <http://www.sciencedirect.com/science/article/pii/S0957417415002158>.
- [20] Gross Charles W., Sohl Jeffrey E. Disaggregation methods to expedite product line forecasting // Journal of Forecasting. — 1990. — Vol. 9, no. 3. — P. 233–254.
- [21] Griliches Zvi, Grunfeld Yehuda. Is Aggregation Necessarily Bad? // Review of Economics and Statistics. — 1960. — Vol. 42, no. 1. — P. 1–13.
- [22] Fogarty Donald W., Blackstone John H., Hoffmann Thomas R. Production & inventory management. — South-Western Publishing Company, 1991.
- [23] Narasimhan Seetharama L., McLeavey Dennis W., Billington Peter. Production planning and inventory control. — Prentice Hall, 1995.
- [24] Fliedner Gene. An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation // Computers & Operations Research. — 1999. — Vol. 26, no. 10. — P. 1133–1149.
- [25] Orcutt Guy H., Watts Harold W., Edwards John B. Data aggregation and information loss // The American Economic Review. — 1968. — Vol. 58, no. 4. — P. 773–787.
- [26] Edwards John B., Orcutt Guy H. Should aggregation prior to estimation be the rule? // The Review of Economics and Statistics. — 1969. — Vol. 51, no. 4. — P. 409–420.
- [27] Schwarzkopf Albert B., Tersine Richard J., Morris John S. Top-down versus bottom-up forecasting strategies // The International Journal Of Production Research. — 1988. — Vol. 26, no. 11. — P. 1833–1843.
- [28] Kinney William R. Predicting earnings: entity versus subentity data // Journal of Accounting Research. — 1971. — Vol. 9, no. 1. — P. 127–136.
- [29] Collins Daniel W. Predicting earnings with sub-entity data: Some further evidence // Journal of Accounting Research. — 1976. — Vol. 14, no. 1. — P. 163–177.

- [30] Dunn D.M., Williams W.H., DeChaine T.L. Aggregate versus subaggregate models in local area forecasting // Journal of the American Statistical Association. — 1976. — Vol. 71, no. 353. — P. 68–71.
- [31] Zellner Arnold, Tobias Justin. A note on aggregation, disaggregation and forecasting performance // Journal of Forecasting. — 2000. — Vol. 19, no. 5. — P. 457–465.
- [32] Dangerfield Byron J., Morris John S. Top-down or bottom-up: Aggregate versus disaggregate extrapolations // International Journal of Forecasting. — 1992. — Vol. 8, no. 2. — P. 233–241.
- [33] Fair Ray C., Shiller Robert J. Comparing information in forecasts from econometric models // The American Economic Review. — 1990. — Vol. 80, no. 3. — P. 375–389.
- [34] Marcellino Massimiliano, Stock James H., Watson Mark W. Macroeconomic forecasting in the euro area: Country specific versus area-wide information // European Economic Review. — 2003. — Vol. 47, no. 1. — P. 1–18.
- [35] Espasa Antoni, Senra Eva, Albacete Rebeca. Forecasting inflation in the European Monetary Union: A disaggregated approach by countries and by sectors // The European Journal of Finance. — 2002. — Vol. 8, no. 4. — P. 402–421.
- [36] Hubrich Kirstin. Forecasting euro area inflation: Does aggregating forecasts by HICP component improve forecast accuracy? // International Journal of Forecasting. — 2005. — Vol. 21, no. 1. — P. 119–136.
- [37] Tiao George C., Guttman Irwin. Forecasting contemporaneous aggregates of multiple time series // Journal of Econometrics. — 1980. — Vol. 12, no. 2. — P. 219–230.
- [38] Kohn Robert. When is an aggregate of a time series efficiently forecast by its past? // Journal of Econometrics. — 1982. — Vol. 18, no. 3. — P. 337–349.
- [39] Shing N.K. A study of bottom-up and top-down forecasting methods. — 1993.
- [40] Fliedner Eugene B., Lawrence Barry. Forecasting system parent group formation: An empirical application of cluster analysis // Journal of Operations Management. — 1995. — Vol. 12, no. 2. — P. 119–130.

- [41] Kahn Kenneth B. Revisiting top-down versus bottom-up forecasting // JOURNAL OF BUSINESS FORECASTING METHODS AND SYSTEMS. — 1998. — Vol. 17, no. 2. — P. 14–19.
- [42] Fliedner Gene. Hierarchical forecasting: issues and use guidelines // Industrial Management & Data Systems. — 2001. — Vol. 101, no. 1. — P. 5–12.
- [43] Forecasting in Hierarchical Environments / Robert Lorenz, Lars Dannecker, Philipp Rösch et al. // Proceedings of the 25th International Conference on Scientific and Statistical Database Management. — SSDBM. — New York, NY, USA : ACM, 2013. — P. 37:1–37:4.
- [44] Efficient Forecasting for Hierarchical Time Series / Lars Dannecker, Robert Lorenz, Philipp Rösch et al. // Proceedings of the 22Nd ACM International Conference on Conference on Information; Knowledge Management. — CIKM '13. — New York, NY, USA : ACM, 2013. — P. 2399–2404.
- [45] Park Mijung, Nassar Marcel. Variational Bayesian inference for forecasting hierarchical time series // Divergence Methods in Probabilistic Inference (DMPI) workshop at International Conference on Machine Learning (ICML). — Beijing, China, 2014.
- [46] Moon Seongmin, Hicks Christian, Simpson Andrew. The development of a hierarchical forecasting method for predicting spare parts demand in the South Korean Navy—A case study // International Journal of Production Economics. — 2012. — Vol. 140, no. 2. — P. 794–802.
- [47] Forecasting the data cube: A model configuration advisor for multi-dimensional data sets / Ulrike Fischer, Christopher Schildt, Claudio Hartmann, Wolfgang Lehner // Data Engineering (ICDE), 2013 IEEE 29th International Conference on / IEEE. — 2013. — P. 853–864.
- [48] Stone Richard, Champernowne David G., Meade James E. The precision of national income estimates // The Review of Economic Studies. — 1942. — Vol. 9, no. 2. — P. 111–125.

- [49] Solomou Solomos, Weale Martin. Balanced estimates of UK GDP 1870–1913 // *Explorations in Economic History*. — 1991. — Vol. 28, no. 1. — P. 54–63.
- [50] Solomou Solomos, Weale Martin. Balanced estimates of national accounts when measurement errors are autocorrelated: the UK, 1920–38 // *Journal of the Royal Statistical Society. Series A (Statistics in Society)*. — 1993. — Vol. 156, no. 1. — P. 89–105.
- [51] Solomou Solomos, Weale Martin. UK national income, 1920–1938: the implications of balanced estimates // *The Economic History Review*. — 1996. — Vol. 49, no. 1. — P. 101–115.
- [52] Weale Martin. Testing linear hypothesis on national account data // *The Review of Economics and Statistics*. — 1985. — Vol. 67, no. 4. — P. 685–689.
- [53] Weale Martin. The reconciliation of values, volumes and prices in the national accounts // *Journal of the Royal Statistical Society. Series A (Statistics in Society)*. — 1988. — Vol. 151, no. 1. — P. 211–221.
- [54] Hyndman Rob J., Athanasopoulos George. Optimally Reconciling Forecasts in a Hierarchy // *Foresight: The International Journal of Applied Forecasting*. — 2014. — Fall. — no. 35. — P. 42–48.
- [55] Van Erven Tim, Cugliari Jairo. Game-theoretically Optimal Reconciliation of Contemporaneous Hierarchical Time Series Forecasts. — 2013. — URL: <http://hal.inria.fr/hal-00920559>.
- [56] Tibshirani Robert. Regression shrinkage and selection via the lasso // *Journal of the Royal Statistical Society. Series B (Methodological)*. — 1996. — Vol. 58, no. 1. — P. 267–288.
- [57] Вычислимые модели и численные методы для анализа тарифной политики железнодорожных грузоперевозок / М.П. Ващенко, А.В. Гасников, Е.Г. Молчанов и др. — М.: ВЦ РАН, 2014.
- [58] Гасников А.В. Заметка об эффективной вычислимости конкурентных равновесий в транспортно-экономических моделях // *Матем. моделирование*. — 2015.

- [59] Bregman Lev M. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming // USSR computational mathematics and mathematical physics. — 1967. — Vol. 7, no. 3. — P. 200–217.
- [60] Gneiting Tilmann. Making and evaluating point forecasts // Journal of the American Statistical Association. — 2011. — Vol. 106, no. 494. — P. 746–762.
- [61] Patton Andrew J. Volatility forecast comparison using imperfect volatility proxies // Journal of Econometrics. — 2011. — Vol. 160, no. 1. — P. 246–256.
- [62] Global Energy Forecasting Competition 2012. — 2012. — URL: <https://www.kaggle.com/c/global-energy-forecasting-competition-2012-load-forecasting>.